

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

#5-7-3-01
S00P1563 US00
Priority
JC886 U.S. PTO
09/748471
12/26/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
る事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
this Office.

願年月日
Date of Application:

1999年12月28日

願番号
Application Number:

平成11年特許願第375766号

願人
Applicant(s):

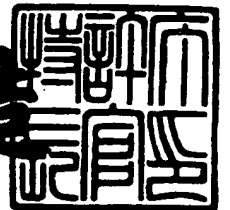
ソニー株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年10月13日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2000-3085029

【書類名】 特許願

【整理番号】 9900847204

【提出日】 平成11年12月28日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/18
G10L 3/00

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 包 洪長

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100082131

【弁理士】

【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 モデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置

【特許請求の範囲】

【請求項 1】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を行うモデル適応装置であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出手段と、

前記所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、前記所定のモデルの適応を行うモデル適応手段と

を備えることを特徴とするモデル適応装置。

【請求項 2】 前記パターン認識は、前記入力データの特徴空間における特徴分布に基づいて行われる

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項 3】 前記モデル適応手段は、前記所定区間における抽出データの時間的位置に対応して値が変化する関数を、前記新鮮度として用いて、前記所定のモデルの適応を行う

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項 4】 前記関数は、時間の経過とともに増加する単調増加関数である

ことを特徴とする請求項 3 に記載のモデル適応装置。

【請求項 5】 前記関数は、線形または非線形の関数である

ことを特徴とする請求項 4 に記載のモデル適応装置。

【請求項 6】 前記関数は、離散値または連続値をとる

ことを特徴とする請求項 4 に記載のモデル適応装置。

【請求項 7】 前記関数は、2 次関数または 3 次以上の高次関数である

ことを特徴とする請求項 4 に記載のモデル適応装置。

【請求項 8】 前記関数は、対数関数である

ことを特徴とする請求項 4 に記載のモデル適応装置。

【請求項 9】 前記入力データは、音声データである

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項 1 0】 前記所定のモデルは、音声区間でない区間におけるノイズを表す音響モデルである

ことを特徴とする請求項 9 に記載のモデル適応装置。

【請求項 1 1】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を行うモデル適応方法であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出ステップと、

前記所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、前記所定のモデルの適応を行うモデル適応ステップと

を備えることを特徴とするモデル適応方法。

【請求項 1 2】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出ステップと、

前記所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、前記所定のモデルの適応を行うモデル適応ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

【請求項 1 3】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識装置であって、

前記入力データの特徴量を抽出する特徴抽出手段と、

前記所定数のモデルを記憶する記憶手段と、

前記入力データの特徴量を、前記所定数のモデルのうちのいずれかに分類する分類手段と、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出手段と、

前記所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、前記所定のモデルの適応を行うモデル適応手段とを備えることを特徴とするパターン認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、モデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置に関し、特に、例えば、音声認識等を行う場合に用いて好適なモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置に関する。

【0002】

【従来の技術】

従来より、ノイズ環境下において発話された単語等を認識する方法が知られており、その代表的な方法としては、PMC(Parallel Model Combination)法、SS/NS S(Spectral Subtraction/Nonlinear Spectral Subtraction)法、SFE(Stochastic Feature Extraction)法等がある。

【0003】

PMC法は、環境ノイズの情報を直接的に音響モデルに取り込んでいるので認識性能は良いが、計算コストが高い（高度な演算を必要とするので、装置の規模が大型化する、処理に要する時間が長い等）。SS/NSS法では、音声データの特徴量を抽出する段階において、環境ノイズが除去される。したがって、PMC法よりも計算コストが低く、現在、多く用いられている方法である。SFE法では、SS/NSS法と同様に、環境ノイズを含む音声信号の特徴量を抽出する段階において、環境ノイズが除去されるが、特徴量として、確率分布で表されるものが抽出される。SFE法では、このように、音声の特徴量が、特徴空間上の分布として抽出される点で、音声の特徴量が、特徴空間上の点として抽出されるSS/NSS法や、PMC法と異なる。

【0004】

上述したいずれの方法においても、音声の特徴量の抽出後は、その特徴量が、予め登録されている複数の単語等に対応する音響モデルのうちのいずれに最も適合するかが判定されて、最も適合する音響モデルに対応する単語が認識結果として出力される。

【 0 0 0 5 】

なお、SFE法については、本件出願人が先に出願した特開平11-133992号（特願平9-300979号）等に、その詳細が記載されている。また、PMC法、SS/NSS法、およびSFE法の性能比較等は、例えば、"H.Pao, H.Honda, K.Minamino, M.Omote, H.Ogawa and N.Iwahashi, Stochastic Feature Extraction for Improving Noise Robustness in Speech Recognition, Proceedings of the 8th Sony Research Forum, SRF98-234, pp.9-14, October 1998", "N.Iwahashi, H.Pao, H.Honda, K.Minamino and M.Omote, Stochastic Features for Noise Robust in Speech Recognition, ICASSP'98 Proceedings, pp.633-636, May, 1998", "N.Iwahashi, H.Pao(presented), H.Honda, K.Minamino and M.Omote, Noise Robust Speech Recognition Using Stochastic Representation of Features, ASJ'98-Spring Proceedings, pp.91-92, March, 1998", "N.iwahashi, H.Pao H.Honda, K.Minamino and M.Omote, Stochastic Represetation of Feature for Noise Robust Speech Recognition, Technical Report of IEICE, pp.19-24, SP97-97(1998-01)等に、その詳細が記載されている。

【 0 0 0 6 】

【発明が解決しようとする課題】

ところで、上述のSFE法等においては、音声認識の段階で環境ノイズが直接的に反映されていない、すなわち、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないので、認識性能が劣る課題があった。

【 0 0 0 7 】

また、環境ノイズの情報が直接的に無音音響モデルに取り込まれていないことに起因して、音声認識が開始された時点から発話が始まるまでの時間が長くなるにつれて認識性能が低下する課題があった。

【 0 0 0 8 】

本発明はこのような状況に鑑みてなされたものであり、環境ノイズの情報を用いて無音音響モデルを補正することにより、音声認識が開始された時から発話が始まるまでの時間が長くなるに伴って認識性能が低下することを抑止することができるようにするものである。

【 0 0 0 9 】

【課題を解決するための手段】

本発明のモデル適応装置は、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、パターン認識に用いる所定のモデルの適応を行うモデル適応手段を備えることを特徴とする。

【 0 0 1 0 】

パターン認識は、入力データの特徴空間における特徴分布に基づいて行われるものとすることができる。

【 0 0 1 1 】

モデル適応手段には、所定区間における抽出データの時間的位置に対応して値が変化する関数を、新鮮度として用いて、所定のモデルの適応を行わせることができる。

【 0 0 1 2 】

関数は、時間の経過とともに増加する単調増加関数とすることができる。

【 0 0 1 3 】

関数は、線形または非線形の関数とすることができる。

【 0 0 1 4 】

関数は、離散値または連続値をとるものとすることができる。

【 0 0 1 5 】

関数は、2次関数または3次以上の高次関数とすることができる。

【 0 0 1 6 】

関数は、対数関数とすることができる。

【 0 0 1 7 】

入力データは、音声データとすることができる。

【 0 0 1 8 】

所定のモデルは、音声区間でない区間におけるノイズを表す音響モデルとすることができる。

【0019】

本発明のモデル適応方法は、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、所定のモデルの適応を行うモデル適応ステップを備えることを特徴とする。

【0020】

本発明の記録媒体は、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、所定のモデルの適応を行うモデル適応ステップを備えるプログラムが記録されていることを特徴とする。

【0021】

本発明のパターン認識装置は、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、所定のモデルの適応を行うモデル適応手段を備えることを特徴とする。

【0022】

本発明のモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置においては、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、所定のモデルの適応が行われる。

【0023】

【発明の実施の形態】

図1は、本発明を適用した音声認識装置の一実施の形態の構成例を示している。この音声認識装置において、マイクロフォン1は、認識対象である発話音声、環境ノイズとともに集音し、フレーム化部2に出力する。フレーム化部2は、マイクロフォン1から入力される音声データを、所定の時間間隔（例えば、10ms）で取り出し、その取り出したデータを、1フレームのデータとして出力する。フレーム化部2が出力する1フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトル a として、ノイズ観測区間抽出部3、および特徴抽出部5に供給される。

【0024】

ここで、以下、適宜、第 t フレームの音声データである観測ベクトルを、 $a(t)$ と表す。

【0025】

ノイズ観測区間抽出部 3 は、フレーム化部 2 から入力されるフレーム単位の音声データを所定の時間 (M フレーム分以上) だけバッファリングし、図 2 に示すように、発話スイッチ 4 がオンとされたタイミング t_b から M フレーム分だけ以前のタイミング t_a までをノイズ観測区間 T_n として、そのノイズ観測区間 T_n における M フレーム分の観測ベクトル a を抽出して、特徴抽出部 5、および無音音響モデル補正部 7 に出力する。

【0026】

発話スイッチ 4 は、ユーザが発話を開始するときにユーザによってオンとされ、発話を終了するときにオフとされる。したがって、発話スイッチ 4 がオンとされたタイミング t_b 以前 (ノイズ観測区間 T_n) の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ 4 がオンとされたタイミング t_b から発話スイッチ 4 がオフとされるタイミング t_d までは、音声認識区間とされて、その音声認識区間の音声データが音声認識の対象とされる。

【0027】

特徴抽出部 5 は、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_n の環境ノイズだけが存在する音声データに基づいて、フレーム化部 2 から入力される、タイミング t_b 以降の音声認識区間の観測ベクトル a から環境ノイズ成分を除去して、その特徴量を抽出する。すなわち、特徴抽出部 5 は、例えば、観測ベクトル a としての真 (環境ノイズが除去された) の音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトル y を算出する。なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【0028】

さらに、特徴抽出部 5 は、観測ベクトル a としての音声データに含まれる音声

を、その特徴量の空間（特徴ベクトル空間）に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ（以下、特徴分布パラメータと記述する） Z を、算出した特徴ベクトル y に基づいて算出し、音声認識部 6 に供給する。

【0029】

図 3 は、図 1 の特徴抽出部 5 の詳細な構成例を示している。フレーム化部 2 から入力される観測ベクトル a は、特徴抽出部 5 において、パワースペクトラム分析部 11 に供給される。パワースペクトラム分析部 11 では、観測ベクトル a が、例えば、FFT（高速フーリエ変換）アルゴリズムによってフーリエ変換され、これにより、音声のパワースペクトラムが、特徴ベクトルとして抽出される。なお、ここでは、1 フレームの音声データとしての観測ベクトル a が、 D 個のコンポーネントからなる特徴ベクトル（ D 次元の特徴ベクトル）に変換されるものとする。

【0030】

ここで、第 t フレームの観測ベクトル $a(t)$ から得られる特徴ベクトルを $y(t)$ と表す。また、特徴ベクトル $y(t)$ のうち、真の音声のスペクトル成分を $x(t)$ と、環境ノイズのスペクトル成分を $u(t)$ と表す。この場合、真の音声のスペクトル成分 $x(t)$ は、次式（1）で表される。

【数 1】

$$x(t) = y(t) - u(t) \quad \dots (1)$$

ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトル $a(t)$ としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定している。

【0031】

一方、ノイズ観測区間抽出部 3 から入力される音声データとしての環境ノイズは、特徴検出部 5 において、ノイズ特性算出部 13 に入力される。ノイズ特性算出部 13 では、ノイズ観測区間 T_n における環境ノイズの特性が求められる。

【0032】

すなわち、ここでは、音声認識区間における環境ノイズのパワースペクトラム

$u(t)$ の分布が、その音声認識区間の直前のノイズ観測区間 T_n における環境ノイズと同一であり、かつ、その分布が正規分布であると仮定して、ノイズ特性算出部 13 において、その正規分布を規定する、環境ノイズの平均値（平均ベクトル）と分散（分散マトリクス）が求められる。

【0033】

なお、平均ベクトル μ' と分散マトリクス Σ' は、次式 (2) にしたがって求めることができる。

【数 2】

$$\begin{aligned}\mu'(i) &= \frac{1}{M} \sum_{t=1}^M y(t)(i) \\ \Sigma'(i, j) &= \frac{1}{M} \sum_{t=1}^M (y(t)(i) - \mu'(i))(y(t)(j) - \mu'(j)) \\ &\dots (2)\end{aligned}$$

ただし、 $\mu'(i)$ は、平均ベクトル μ' の i 番目のコンポーネントを表す ($i = 1, 2, \dots, D$)。また、 $y(t)(i)$ は、第 t フレームの特徴ベクトルの i 番目のコンポーネントを表す。さらに、 $\Sigma'(i, j)$ は、分散マトリクス Σ' の、第 i 行、第 j 列のコンポーネントを表す ($j = 1, 2, \dots, D$)。

【0034】

ここで、計算量の低減のために、環境ノイズについては、特徴ベクトル y の各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクス Σ' は、対角成分以外は 0 となる。

【数 3】

$$\Sigma'(i, j) = 0, i \neq j \quad \dots (3)$$

【0035】

ノイズ特性算出部 13 では、以上のようにして、環境ノイズの特性としての、正規分布を規定する平均ベクトル μ' および平均値 Σ' が求められ、特徴分布パラメータ算出部 12 に供給される。

【0036】

一方、パワースペクトラム分析部 1 1 の出力、すなわち、環境ノイズを含む発話音声の特徴ベクトル y は、特徴分布パラメータ算出部 1 2 に供給される。特徴分布パラメータ算出部 1 2 では、パワースペクトラム分析部 1 1 からの特徴ベクトル y 、およびノイズ特性算出部 1 3 からの環境ノイズの特性に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータが算出される。

【 0 0 3 7 】

すなわち、特徴分布パラメータ算出部 1 2 では、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトル ξ と分散マトリクス Ψ が、特徴分布パラメータとして、次式（4）乃至（7）にしたがって計算される。

【数 4】

$$\begin{aligned}
 \xi(t)(i) &= E[x(t)(i)] \\
 &= E[y(t)(i) - u(t)(i)] \\
 &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i)) \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\
 &= \frac{y(t)(i) \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) - \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\
 &= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (4)
 \end{aligned}$$

【数 5】

 $i=j$ のとき

$$\begin{aligned}
 \Psi(t)(i, j) &= V[x(t)(i)] \\
 &= E[(x(t)(i))^2] - (E[x(t)(i)])^2 \\
 &= E[(x(t)(i))^2] - (\xi(t)(i))^2
 \end{aligned}$$

 $i \neq j$ のとき

$$\Psi(t)(i, j) = 0 \quad \dots (5)$$

【数 6】

$$E[(x(t)(i))^2] = E[(y(t)(i) - u(t)(i))^2]$$

$$\begin{aligned}
 &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i))^2 \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\
 &= \frac{1}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \times \left\{ (y(t)(i))^2 \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) \right. \\
 &\quad \left. - 2y(t)(i) \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i) \right. \\
 &\quad \left. + \int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i) \right\} \\
 &= (y(t)(i))^2 - 2y(t)(i) \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\
 &\quad + \frac{\int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (6)
 \end{aligned}$$

【数 7】

$$P(u(t)(i)) = \frac{1}{\sqrt{2\pi \Sigma'(i, i)}} e^{-\frac{1}{2\Sigma'(i, i)}(u(t)(i) - \mu'(i))^2} \dots (7)$$

【0038】

ここで、 $\xi(t)(i)$ は、第 t フレームにおける平均ベクトル $\xi(t)$ の i 番目のコンポーネントを表す。また、 $E[\]$ は、 $[\]$ 内の平均値を意味する。 $x(t)(i)$ は、第 t フレームにおける真の音声のパワースペクトラム $x(t)$ の i 番目のコンポーネントを表す。さらに、 $u(t)(i)$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントを表し、 $P(u(t)(i))$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントが $u(t)(i)$ である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、 $P(u(t)(i))$ は、式 (7) に示したように表される。

【0039】

また、 $\Psi(t)(i, j)$ は、第 t フレームにおける分散 $\Psi(t)$ の、第 i 行、第 j 列のコンポーネントを表す。さらに、 $V[\]$ は、 $[\]$ 内の分散を表す。

【0040】

特徴分布パラメータ算出部 12 では、以上のようにして、各フレームごとに、平均ベクトル ξ および分散マトリクス Ψ が、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求められる。

【0041】

その後、音声認識区間の各フレームにおいて求めた特徴分布パラメータは、音声認識部 6 に出力される。すなわち、いま、音声認識区間が T フレームであったとし、その T フレームそれぞれにおいて求められた特徴分布パラメータを、 $z(t) = \{\xi(t), \Psi(t)\}$ ($t = 1, 2, \dots, T$) と表すと、特徴分布パラメータ算出部 12 は、特徴分布パラメータ（系列） $Z = \{z(1), z(2), \dots, z(T)\}$ を、音声認識部 6 に供給する。

【0042】

図1に戻り、音声認識部6は、特徴抽出部5から入力される特徴分布パラメータ Z を、所定数 K の音響モデルと1個の無音音響モデルのうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部6は、例えば、無音区間に対応する識別関数（特徴パラメータ Z が無音音響モデルに分類されるかを識別するための関数）と、所定数 K の単語それぞれに対応する識別関数（特徴パラメータ Z がいずれの音響モデルに分類されるかを識別するための関数）とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部5からの特徴分布パラメータ Z を引数として計算する。そして、その関数値（いわゆるスコア）が最大である音響モデル（単語、または無音（ノイズ））が認識結果として出力される。

【0043】

即ち、図4は、図1の音声認識部6の詳細な構成例を示している。特徴抽出部5の特徴分布パラメータ算出部12から入力される特徴分布パラメータ Z は、識別関数演算部21-1乃至21- k 、および識別関数演算部21- s に供給される。識別関数演算部21- k （ $k=1, 2, \dots, K$ ）は、 K 個の音響モデルのうちの k 番目に対応する単語を識別するための識別関数 $G_k(Z)$ を記憶しており、特徴抽出部5からの特徴分布パラメータ Z を引数として、識別関数 $G_k(Z)$ を演算する。識別関数演算部21- s は、無音音響モデルに対応する無音区間を識別するための識別関数 $G_s(Z)$ を記憶しており、特徴抽出部5からの特徴分布パラメータ Z を引数として、識別関数 $G_s(Z)$ を演算する。

【0044】

なお、音声認識部6では、例えば、HMM(Hidden Markov Model)法を用いて、クラスとしての単語または無音の識別（認識）が行われる。

【0045】

HMM法について、図5を参照して説明する。同図において、HMMは、 H 個の状態 q_1 乃至 q_H を有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態 q_1 とされ、最終状態は、最も右の状態 q_H とされており、最終状態 q_H からの状態遷移は禁止されている。

。このように、自身よりも左にある状態への遷移のないモデルは、left-to-rightモデルと呼ばれ、音声認識では、一般に、left-to-rightモデルが用いられる。

いま、HMMのkクラスを識別するためのモデルを、kクラスモデルというすると、kクラスモデルは、例えば、最初に状態 q_h にいる確率（初期状態確率） $\pi_k(q_h)$ 、ある時刻（フレーム）tにおいて、状態 q_i にいて、次の時刻t+1において、状態 q_j に状態遷移する確率（遷移確率） $a_k(q_i, q_j)$ 、および状態 q_i から状態遷移が生じるときに、その状態 q_i が、特徴ベクトルOを出力する確率（出力確率） $b_k(q_i)(O)$ によって規定される（ $h=1, 2, \dots, H$ ）。

【0046】

そして、ある特徴ベクトル系列 O_1, O_2, \dots が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【0047】

ここでは、この観測確率が、識別関数 $G_k(Z)$ によって求められる。すなわち、識別関数 $G_k(Z)$ は、特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ が観測される確率を求めるものとして、次式（8）で与えられる。

【数8】

$$g_k(Z) = \max_{q_1, q_2, \dots, q_T} \pi_k(q_1) \cdot b'_k(q_1)(z_1) \cdot a_k(q_1, q_2) \cdot b'_k(q_2)(z_2) \\ \cdot \dots \cdot a_k(q_{T-1}, q_T) \cdot b'_k(q_T)(z_T) \\ \dots (8)$$

【0048】

ここで、 $b'_k(q_i)(z_j)$ は、出力が z_j で表される分布であるときの出力確率を表す。状態遷移時に各特徴ベクトルを出力する確率である出力確率 $b_k(s)(O_t)$ には、ここでは、例えば、特徴ベクトル空間上のコンポーネントに相関がないものとして、正規分布関数が用いられている。この場合、入力が z_t

で表される分布であるとき、出力確率 $b_k'(s)(z_t)$ は、平均ベクトル $\mu_k(s)$ と分散マトリクス $\Sigma_k(s)$ とによって規定される確率密度関数 $P_k^m(s)(x)$ 、および第 t フレームの特徴ベクトル（ここでは、パワースペクトラム） x の分布を表す確率密度関数 $P^f(t)(x)$ を用いて、次式（9）により求めることができる。

【数 9】

$$\begin{aligned} b_k'(s)(z_t) &= \int P^f(t)(x) P_k^m(s)(x) dx \\ &= \prod_{i=1}^D P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) \\ &\quad k=1, 2, \dots, K: s=q_1, q_2, \dots, q_T: T=1, 2, \dots, T \\ &\quad \dots (9) \end{aligned}$$

ただし、式（9）における積分の積分区間は、 D 次元の特徴ベクトル空間（ここでは、パワースペクトラム空間）の全体である。

【0 0 4 9】

また、式（9）において、 $P(s)(i)(\xi(t)(i), \Psi(t)(i, i))$ は、次式（10）で表される。

【数 1 0】

$$\begin{aligned} P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) &= \frac{1}{\sqrt{2\pi(\Sigma_k(s)(i, i) + \Psi(t)(i, i))}} e^{-\frac{(\mu_k(s)(i) - \xi(t)(i))^2}{2(\Sigma_k(s)(i, i) + \Psi(t)(i, i))}} \\ &\quad \dots (10) \end{aligned}$$

ただし、 $\mu_k(s)(i)$ は、平均ベクトル $\mu_k(s)$ の i 番目のコンポーネントを、 $\Sigma_k(s)(i, i)$ は、分散マトリクス $\Sigma_k(s)$ の、第 i 行第 i 列のコンポーネントを、それぞれ表す。そして、 k クラスモデルの出力確率は、これらによって規定される。

【0 0 5 0】

なお、HMMは、上述したように、初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i, q_j)$ 、および出力確率 $b_k(q_i)(O)$ によって規定されるが、これらは、

学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めることとする。

【0051】

ここで、HMMとして、図5に示したものをを用いる場合には、常に、最も左の状態 q_1 から遷移が始まるので、状態 q_1 に対応する初期状態確率だけが1とされ、他の状態に対応する初期状態確率はすべて0とされる。また、出力確率は、式(9)、(10)から明らかなように、 $\Psi(t)(i, i)$ を0とすると、特徴ベクトルの分散を考慮しない場合の連続HMMにおける出力確率に一致する。

【0052】

なお、HMMの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【0053】

図4に戻る。識別関数演算部 21-k ($k=1, 2, \dots, K$) は、 k クラスモデルについて、あらかじめ学習により求められている初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i, q_j)$ 、および出力確率 $b_k(q_i)(O)$ によって規定される式(8)の識別関数 $G_k(Z)$ を記憶しており、特徴抽出部 2 からの特徴分布パラメータ Z を引数として、識別関数 $G_k(Z)$ を演算し、その関数値(上述した観測確率) $G_k(Z)$ を、決定部 22 に出力する。識別関数演算部 21-s は、無音音響モデル補正部 7 から供給される初期状態確率 $\pi_s(q_h)$ 、遷移確率 $a_s(q_i, q_j)$ 、および出力確率 $b_s(q_i)(O)$ によって規定される、式(8)の識別関数 $G_k(Z)$ と同様の識別関数 $G_s(Z)$ を記憶しており、特徴抽出部 2 からの特徴分布パラメータ Z を引数として、識別関数 $G_s(Z)$ を演算し、その関数値(上述した観測確率) $G_s(Z)$ を、決定部 22 に出力する。

【0054】

決定部 22 では、識別関数演算部 21-1 乃至 21-k、および識別関数演算部 21-s それぞれからの関数値 $G_k(Z)$ (ここでは、関数値 $G_s(Z)$ を含むものとする) に対して、例えば、次式(11)に示す決定規則を用いて、特徴分布パラメータ Z 、すなわち、入力された音声に属するクラス(音響モデル)が識別される。

【数 1 1】

$$C(Z) = C_k, \text{ if } G_k(Z) = \max_i \{G_i(Z)\} \quad \dots (11)$$

ただし、 $C(Z)$ は、特徴分布パラメータ Z が属するクラスを識別する識別操作（処理）を行う関数を表す。また、式（1 1）の第 2 式の右辺における \max は、それに続く関数値 $G_i(Z)$ （ただし、ここでは、 $i = s, 1, 2, \dots, K$ ）の最大値を表す。

【0 0 5 5】

決定部 2 2 は、式（1 1）にしたがって、クラスを決定すると、それを、入力された音声の認識結果として出力する。

【0 0 5 6】

図 1 に戻り、無音音響モデル補正部 7 は、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_n の音声データとしての環境ノイズに基づいて、音声認識部 6 に記憶されている無音音響モデルに対応する識別関数 $G_s(Z)$ を生成して音声認識部 6 に供給する。

【0 0 5 7】

具体的には、無音音響モデル補正部 7 では、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_n の音声データ（環境ノイズ）の M 個のフレームの各フレームについて、特徴ベクトル X が観測され、それらの特徴分布が生成される。

【数 1 2】

$$\{F_1(X), F_2(X), \dots, F_M(X)\} \quad \dots (12)$$

なお、特徴分布 $\{F_i(X), i = 1, 2, \dots, M\}$ は、確率密度関数 (Probabilistic Density Function) であり、以下、無音特徴分布 PDF とも記述する。

【0 0 5 8】

次に、無音特徴分布 PDF を、次式（1 3）に従い、無音音響モデルに対応する確率分布 $F_s(X)$ に写像する。

【数 1 3】

$$F_s(X) = V(F_1(X), F_2(X), \dots, F_M(X)) \quad \dots (13)$$

ただし、 V は無音特徴分布 PDF $\{F_i(X), i = 1, 2, \dots, M\}$ を無音音

響モデル $F_s(X)$ に写像する補正関数（写像関数）である。

【0059】

この写像は、無音特徴分布PDFの記述によって様々な方法が考えられる。例えば、

【数14】

$$F_s(X) = \sum_{i=1}^M \beta_i(F_1(X), F_2(X), \dots, F_M(X), M) \cdot F_i(X) \quad \dots (14)$$

$$= \sum_{i=1}^M \beta_i \cdot F_i(X) \quad \dots (15)$$

ただし、 $\beta_i(F_1(X), F_2(X), \dots, F_M(X), M)$ は、各無音特徴分布に対する重み関数であり、以下、 β_i と記述する。なお、重み関数 β_i は、次式（16）の条件を満足するものである。

【数15】

$$\sum_{i=1}^M \beta_i(F_1(X), F_2(X), \dots, F_M(X), M) = \sum_{i=1}^M \beta_i \equiv 1 \quad \dots (16)$$

【0060】

ここで、無音音響モデルの確率分布 $F_s(X)$ が正規分布であると仮定し、また、各フレームの特徴ベクトルを構成するコンポーネントが無相関であると仮定すれば、無音特徴分布PDF $\{F_i(X), i = 1, 2, \dots, M\}$ の共分散行列 Σ_i は対角線行列となる。ただし、この仮定的前提条件として、無音音響モデルの共分散行列も対角線行列であることが必要である。したがって、各フレームの特徴ベクトルを構成するコンポーネントが無相関であれば、無音特徴分布PDF $\{F_i(X), i = 1, 2, \dots, M\}$ は、各コンポーネントに対応する平均と分散を持つ正規分布 $G(E_i, \Sigma_i)$ となる。 E_i は $F_i(X)$ の平均値（以下、適宜、期待値ともいう）であり、 Σ_i は $F_i(X)$ の共分散行列である。

【0061】

さらに、ノイズ観測区間 T_n の M 個のフレームに対応する無音特徴分布の平均を μ_i 、分散を σ_i^2 と表すことにすれば、無音特徴分布の確率密度関数は、正規分布 $G(\mu_i, \sigma_i^2)$ ($i = 1, 2, \dots, M$) で表すことができる。以上の

仮定をにより、各フレームに対応する平均 μ_i 、および分散 σ_i^2 を用い、以下に示す様々な方法によって、無音音響モデル $F_s(X)$ を近似する正規分布 $G(\mu_s, \sigma_s^2)$ （上述した $G_s(Z)$ に相当する）を演算することができる。

【0062】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第1の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i=1, 2, \dots, M\}$ を用い、次式(17)に示すように、全ての μ_i の平均を無音音響モデルの平均値 μ_s とし、次式(18)に示すように、全ての σ_i^2 の平均を無音音響モデルの分散 σ_s^2 とする方法である。

【数16】

$$\mu_s = \frac{a}{M} \sum_{i=1}^M \mu_i \quad \dots (17)$$

$$\sigma_s^2 = \frac{b}{M} \sum_{i=1}^M \sigma_i^2 \quad \dots (18)$$

ここで、 a および b は、シミュレーションにより最適な値が決定される係数である。

【0063】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第2の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i=1, 2, \dots, M\}$ の期待値 μ_i のだけを用い、次式(19)、(20)に従って、無音音響モデルの平均値 μ_s と、分散 σ_s^2 を演算する方法である。

【数17】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^M \mu_i \quad \dots (19)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M} \cdot \sum_{i=1}^M \mu_i^2 - \mu_s^2 \right) \quad \dots (20)$$

ここで、 a および b は、シミュレーションにより最適な値が決定される係数である。

【0064】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第3の方法は、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i = 1, 2, \dots, M\}$ の組み合わせによって、無音音響モデルの平均値 μ_s と、分散 σ_s^2 を演算する方法である。

【0065】

この方法においては、各無音特徴分布 $G(\mu_i, \sigma_i^2)$ の確率統計量を X_i とする。

【数18】

$$\{X_1, X_2, \dots, X_M\} \quad \dots (21)$$

【0066】

ここで、無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ の確率統計量を X_s とすれば、確率統計量 X_s は、次式(22)に示すように、確率統計量 X_i と重み関数 β_i の線形結合で表すことができる。なお、重み関数 β_i は式(16)の条件を満足している。

【数19】

$$X_s = \sum_{i=1}^M \beta_i \cdot X_i \quad \dots (22)$$

【0067】

そして、無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ は、次式(23)に示すように表される。

【数20】

$$G(\mu_s, \sigma_s^2) = G\left(\sum_{i=1}^M \beta_i \mu_i, \sum_{i=1}^M \beta_i^2 \sigma_i^2\right) \quad \dots (23)$$

【0068】

なお、式(23)において、重み関数 β_i は、一般には、例えば、 $1/M$ とすることができ、この場合、式(23)の平均値 μ_s と分散 σ_s^2 は、例えば、次式で示すように、所定の係数を用いて求められる。

【数 2 1】

$$\mu_s = \frac{a}{M} \cdot \sum_{i=1}^M \mu_i \quad \dots (24)$$

$$\sigma_s^2 = \frac{b}{M^2} \cdot \sum_{i=1}^M \sigma_i^2 \quad \dots (25)$$

ここで、a および b は、シミュレーションにより最適な値が決定される係数である。

【0 0 6 9】

無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を演算する第 4 の方法では、無音特徴分布 $\{G(\mu_i, \sigma_i^2), i = 1, 2, \dots, M\}$ の確率統計量 X_i に対応する統計母集団 $\Omega_i = \{f_{i,j}\}$ を仮定する。ここで、

【数 2 2】

$\{N_i \equiv N; i = 1, 2, \dots, M\}$

とすれば、平均値 μ_i は、次式 (2 6) によって得ることができ、分散 σ_i^2 は、次式 (2 8) によって得ることができる。

【数 2 3】

$$\mu_i = \frac{1}{N} \sum_{j=1}^M f_{i,j} \quad \dots (26)$$

$$\sigma_i^2 = \frac{1}{N} \sum_{j=1}^M (f_{i,j}^2 - \mu_j^2) \quad \dots (27)$$

$$= \frac{1}{N} \sum_{j=1}^M f_{i,j}^2 - \mu_j^2 \quad \dots (28)$$

【0 0 7 0】

式 (2 8) を変形すれば、次式 (2 9) の関係が成立する。

【数 2 4】

$$\frac{1}{N} \sum_{j=1}^M f_{i,j}^2 = \sigma_i^2 + \mu_i^2 \quad \dots (29)$$

【0 0 7 1】

ここで、統計母集団の和 Ω

【数 2 5】

$$\Omega = \bigcup_{i=1}^M \Omega_i$$

を考慮すれば、式 (2 6) から次式 (3 0) , (3 1) が導かれ、式 (2 9) から次式 (3 2) 乃至 (3 4) が導かれる。

【数 2 6】

$$\mu_s = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f_{i,j} \quad \dots (30)$$

$$= \frac{1}{M} \sum_{i=1}^M \mu_i \quad \dots (31)$$

$$\sigma_s^2 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{i,j} - \mu_s)^2 \quad \dots (32)$$

$$= \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f_{i,j}^2 - \mu_s^2 \quad \dots (33)$$

$$= \frac{1}{M} \sum_{i=1}^M (\sigma_i^2 + \mu_i^2) - \mu_s^2 \quad \dots (34)$$

【0 0 7 2】

なお、実際には、式 (3 1) と式 (3 4) には係数が乗算されて用いられる。

【数 2 7】

$$\mu_s = \frac{a}{M} \sum_{i=1}^M \mu_i \quad \dots (35)$$

$$\sigma_s^2 = b \cdot \left(\frac{1}{M} \sum_{i=1}^M (\sigma_i^2 + \mu_i^2) - \mu_s^2 \right) \quad \dots (36)$$

ここで、a および b は、シミュレーションにより最適な値が決定される係数である。

【0 0 7 3】

また、次式 (3 7) に示すように、分散 σ_s^2 に対してだけ、係数を乗算するようにしてもよい。

【数 2 8】

$$\sigma_s^2 = \frac{b}{M} \sum_{i=1}^M \sigma_i^2 + \frac{1}{M} \sum_{i=1}^M \mu_i^2 - \mu_s^2 \quad \dots (37)$$

【0 0 7 4】

次に、図 1 の音声認識装置の動作について説明する。

【0 0 7 5】

フレーム化部 2 には、マイクロフォン 1 で集音された音声データ（環境ノイズを含む認識対象の発話音声）が入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトル a として、ノイズ観測区間抽出部 3、および特徴抽出部 5 に順次供給される。ノイズ観測区間抽出部 3 では、発話スイッチ 4 がオンとされたタイミング t_b 以前のノイズ観測区間 T_n の音声データ（環境ノイズ）が抽出されて、特徴抽出部 5 および無音音響モデル補正部 7 に供給される。

【0 0 7 6】

無音音響モデル補正部 7 では、ノイズ観測区間 T_n の音声データとしての環境ノイズに基づいて、上述した第 1 乃至第 4 の方法のうちのいずれかによって、無音音響モデルの更新（適応）が行われ、音声認識部 6 に供給される。音声認識部 6 では、無音音響モデル補正部 7 から供給される無音音響モデルとしての識別関数によって、それまで記憶されていた無音音響モデルに対応する識別関数が更新される。即ち、無音音響モデルの適応が行われる。

【0 0 7 7】

一方、特徴抽出部 5 では、フレーム化部 2 からの観測ベクトル a としての音声データが音響分析され、その特徴ベクトル y が求められる。さらに、特徴抽出部 5 では、求められた特徴ベクトル y に基づいて、特徴ベクトル空間における分布を表す特徴分布パラメータ Z が算出され、音声認識部 6 に供給される。音声認識部 6 では、特徴抽出部 5 からの特徴分布パラメータを用いて、無音および所定数 K の単語それぞれに対応する音響モデルの識別関数の値が演算され、その関数値が最大となる音響モデルが、音声の認識結果として出力される。

【0 0 7 8】

以上のように、観測ベクトル a としての音声データが、その特徴量の空間である特徴ベクトル空間における分布を表す特徴分布パラメータ Z に変換されるので、その特徴分布パラメータは、音声データに含まれるノイズの分布特性を考慮したものとなっており、また、無音を識別（認識）するための無音音響モデルに対応する識別関数が、発話直前のノイズ観測区間 T_n の音声データに基づいて更新されているので、音声認識率を大きく向上させることが可能となる。

【0079】

次に、図6は、発話スイッチ4がオンとされてから発話が始まるまでの無音区間 T_s （図2）を変化させたときの音声認識率の変化を測定した実験（シミュレーション）の結果を示している。

【0080】

なお、図6において、曲線aは無音音響モデルを補正しない（無音音響モデルの適応を行わない）従来の方法による結果を示しており、曲線bは第1の方法による結果を示しており、曲線cは第2の方法による結果を示しており、曲線dは第3の方法による結果を示しており、曲線eは、第4の方法による結果を示している。

【0081】

実験の条件は、以下の通りである。認識に用いた音声データは、高速道路を走行中の車内で集音されたものである。ノイズ観測区間 T_n は、20フレームで約0.2秒である。無音区間 T_s は、0.05秒、0.1秒、0.2秒、0.3秒、0.5秒とした。音声データの特徴抽出においては、MFCC(Mel-Frequency Cepstral Coefficients)ドメインで分析を実施した（MFCC分析により、特徴量を得た）。認識の対象とする音声の発話者は、男女4人ずつ計8人であり、一人当たり303個の単語を、個別に発話してもらった。認識を行った単語数は、日本語の5000単語である。音響モデルは、HMMであり、学習用に用意した音声データを用いて予め学習を行った。音声認識においては、Viterbiサーチ法を用いて、そのビーム幅を3000とした。

【0082】

なお、第1、第2、および第4の方法においては、係数 a を1.0とし、係数

b を 0. 1 とした。第 3 の方法においては、係数 a を 1. 0 とし、係数 b を 1. 0 とした。

【 0 0 8 3 】

図 6 から明らかなように、従来の方法（曲線 a）では、無音区間 T_s が長くなるのに伴って音声認識率が著しく低下しているが、本発明の第 1 乃至 4 の方法（曲線 b 乃至 e）では、無音区間 T_s が長くなっても、音声認識率は、わずかしかな低下しない。すなわち、本発明によれば、無音区間 T_s が変化しても、音声認識率はある程度のレベルを維持することが可能である。

【 0 0 8 4 】

なお、上述の第 1 乃至第 4 のいずれの方法においても、無音音響モデルの正規分布 $G(\mu_s, \sigma_s^2)$ を規定する平均値 μ_s は、無音特徴分布 $G(\mu_i, \sigma_i^2)$ の平均値 μ_i の平均値となる。従って、例えば、いま、無音特徴分布 $G(\mu_i, \sigma_i^2)$ の平均値 μ_i の平均値を、 μ と表すとともに、第 1 乃至第 4 の方法によって求められる無音音響モデルの正規分布を、それぞれ、 $G_{s1}(\mu, \sigma_{s1}^2)$ 、 $G_{s2}(\mu, \sigma_{s2}^2)$ 、 $G_{s3}(\mu, \sigma_{s3}^2)$ 、 $G_{s4}(\mu, \sigma_{s4}^2)$ と表すと、これらは、図 7 に示すように、特徴空間において、平均値 μ を中心（重心）とする分布となる。

【 0 0 8 5 】

ところで、無音特徴分布 $G(\mu_i, \sigma_i^2)$ に基づく、上述の第 1 乃至第 4 の方法による無音音響モデルの適応は、写像 V を用いて、次の式（38）で定義することができる。なお、以下、適宜、 $G(\mu_i, \sigma_i^2)$ を G_i と、 $G(\mu_s, \sigma_s^2)$ を G_s と、それぞれ記述する。

【 0 0 8 6 】

【数 2 9】

$$G_s(\cdot) = V(G_1, G_2, \dots, G_i, \dots) \quad \dots (38)$$

【 0 0 8 7 】

また、ここでは、無音音響モデル G として、正規分布を仮定しており、正規分布は、平均値と分散で規定されるから、無音音響モデル G の正規分布を規定する平均値と分散を、上述のように、 μ_s と σ_s^2 で表せば、式（38）の定義は、平

均値と分散の写像 V_{μ} と V_{σ^2} とをそれぞれ用いて、式 (39) および (40) で表すこともできる。

【0088】

【数30】

$$\mu_s = V_{\mu}(G_1, G_2, \dots) \quad \dots (39)$$

$$\sigma_s^2 = V_{\sigma^2}(G_1, G_2, \dots) \quad \dots (40)$$

【0089】

上述の写像 V (V_{μ} および V_{σ^2}) で表される第1乃至第4の方法では、ノイズ観測区間 T_n (図2) における M フレームそれぞれから得られる時系列の無音特徴分布 G_1, G_2, \dots, G_M を平等に取り扱っている。

【0090】

しかしながら、音声認識区間における環境ノイズは、厳密には、音声認識区間の直前のノイズ観測区間 T_n における環境ノイズと同一ではなく、さらに、一般には、ノイズ観測区間 T_n における環境ノイズは、音声認識区間 (の開始時刻 t_c) から離れるほど、音声認識区間における環境ノイズとは異なるものとなると推測される。

【0091】

従って、ノイズ観測区間 T_n (図2) における M フレームそれぞれから得られる時系列の無音特徴分布 G_1, G_2, \dots, G_M は、平等に扱うのではなく、音声認識区間に近いものほど重みをおいて扱うべきであり (音声認識区間から遠いものほど重みをおかずに扱うべきであり)、そのようにすることで、音声認識精度をより向上させる無音音響モデルの適応 (補正および更新) が可能となる。

【0092】

そこで、ノイズ観測区間 T_n において得られる無音特徴分布 G_1, G_2, \dots, G_M について、その新しさ (ここでは、音声認識区間への近さに相当する) を表す新鮮度を導入することとし、この新鮮度を考慮して、無音音響モデルの適応を行う方法について説明する。

【0093】

図8は、新鮮度を考慮して、無音音響モデルの適応を行う、図1の無音音響モ

デル補正部 7 の構成例を示している。

【 0 0 9 4 】

新鮮度関数記憶部 3 1 は、上述したような新鮮度を表す関数である新鮮度関数（を規定するパラメータ）を記憶している。

【 0 0 9 5 】

補正部 3 2 には、ノイズ観測区間抽出部 3 が出力する、ノイズ観測区間 T_n における音声データ（ノイズ）としての観測ベクトルの系列（ここでは、 M フレームの音声データ）が入力されるようになっており、補正部 3 2 は、この観測ベクトルから、無音特徴分布 G_1, G_2, \dots, G_M を得て、これらと、新鮮度関数記憶部 3 1 に記憶されている新鮮度関数に基づいて、無音音響モデルの適応を行う。

【 0 0 9 6 】

ここで、無音特徴分布 G_1, G_2, \dots, G_M は、ノイズ観測区間 T_n における M フレームそれぞれで観測される離散値であり、無音音響モデル補正部 7 が、離散値を処理するシステムであれば、離散値である無音特徴分布 G_1, G_2, \dots, G_M をそのまま用いることができる。しかしながら、無音音響モデル補正部 7 が、連続値を処理するシステムである場合には、例えば、図 9 に示すように、離散値である無音特徴分布 G_1, G_2, \dots, G_M を、連続変換器で連続値に変換してから、無音音響モデル補正部 7 で処理する必要がある。離散値を連続値に変換する方法としては、例えば、スプライン関数(Spline Function)によって近似を行う方法がある。

【 0 0 9 7 】

なお、離散値とは、ある有限の観測区間において、離散的な時刻で観測される有限個の観測値であり、連続値とは、ある有限（または無限）の観測区間の任意の時刻で観測される無限個の観測値であり、ある関数によって表現される。

【 0 0 9 8 】

無音音響モデルの適応に用いる無音特徴分布が離散値である場合には、新鮮度関数も離散値の関数となり、無音特徴分布が連続値である場合には、新鮮度関数も連続値の関数となる。

【0 0 9 9】

次に、新鮮度関数、およびそれを用いた無音音響モデルの適応について、新鮮度関数が離散値である場合と、連続値である場合とに分けて説明する。

【0 1 0 0】

まず、新鮮度関数 $F(x)$ は、例えば、式 (4 1) 乃至 (4 3) に示すように定義することができる。

【0 1 0 1】

【数 3 1】

$$F(x) = 0 \quad \text{if } x \notin \Omega_{\text{obs}} \quad \dots (41)$$

$$F(x_2) \geq F(x_1) \quad \text{if } x_2 \geq x_1 \quad \dots (42)$$

$$\int_{\Omega_{\text{obs}}} F(x) dx \equiv 1 \quad \dots (43)$$

但し、 Ω_{obs} は、無音特徴分布の観測区間を表し、本実施の形態では、ノイズ観測区間 T_n に相当する。

【0 1 0 2】

式 (4 1) により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} 以外では 0 となる。また、式 (4 2) により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} において、時間の経過とともに増加するか、または変化しない関数（本明細書において、単調増加関数という）であり、従って、新鮮度関数 $F(x)$ は、基本的に、音声認識区間（図 2）に近づくほど、大きな値となる。さらに、式 (4 3) により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} に亘って積分した場合に、その積分値が 1 となる関数である。式 (4 1) 乃至 (4 3) から、新鮮度関数 $F(x)$ は、例えば、図 1 0 に示すようになる。

【0 1 0 3】

ここで、本実施の形態では、新鮮度関数 $F(x)$ は、後述するように、無音特徴分布に乗算する乗数として用いられる。従って、新鮮度関数 $F(x)$ は、その値が正または負のときには、それが乗数として乗算される無音特徴分布に対する重みとして作用する。また、新鮮度関数 $F(x)$ は、その値が 0 のときは、それが乗数として乗算される無音特徴分布を無効とし、無音音響モデルの適応に影響

を与えないように作用する。

【0 1 0 4】

図 8 の補正部 3 2 では、以上のような新鮮度関数 $F(x)$ と、無音特徴分布 G_1, G_2, \dots, G_M とを用いて、基本的には、式 (4 4) にしたがって、適応後の無音音響モデル G_s が求められる。

【0 1 0 5】

【数 3 2】

$$\begin{aligned} G_s &= V(G_1, \dots, G_M) \\ &= \sum_{x=1}^M F(x) \cdot G_x \end{aligned} \quad \dots (44)$$

【0 1 0 6】

式 (4 4) によれば、無音特徴分布が、音声認識区間に近いもののほど重みをおいて扱われて、無音音響モデルの適応が行われ、その結果、音声認識精度をより向上させることが可能となる。

【0 1 0 7】

次に、新鮮度関数 $F(x)$ の具体例と、それを用いた無音音響モデルの適応について説明する。なお、以下では、無音特徴分布の観測区間 Ω_{obs} (本実施の形態では、ノイズ観測区間 T_n) を、 x が 0 から x_M までの区間とする。また、新鮮度関数 $F(x)$ の関数値としては、観測区間 Ω_{obs} のみの値を考えることとする (式 (4 1) に示したように、新鮮度関数 $F(x)$ の関数値は、観測区間 Ω_{obs} 以外では 0 であるので、以下では、その点については言及しない)。

【0 1 0 8】

新鮮度関数 $F(x)$ としては、例えば、線形の関数を用いることができ、関数値として連続値をとる場合には、新鮮度関数 $F(x)$ は、例えば、式 (4 5) で表される。

【0 1 0 9】

【数 3 3】

$$F(x) = a \cdot x \quad \dots (45)$$

【0 1 1 0】

式(45)における α は、所定の定数であり、この定数 α は、式(43)の新鮮度関数の定義から、 $2/x_M^2$ となる。従って、式(45)の新鮮度関数 $F(x)$ は、式(46)で表されることになる。

【0111】

【数34】

$$F(x) = \frac{2}{x_M^2} \cdot x \quad \dots (46)$$

【0112】

ここで、式(46)で表される新鮮度関数 $F(x)$ を、図11に示す。

【0113】

この場合、適応後の無音音響モデル G_s は、式(47)にしたがって求められる。

【0114】

【数35】

$$G_s = \frac{2}{x_M^2} \cdot \int_0^{x_M} x \cdot G_x(\mu_x, \sigma_x^2) dx \quad \dots (47)$$

なお、 $G_x(\mu_x, \sigma_x^2)$ は、時刻 x における無音特徴分布を表し、 μ_x と σ_x^2 は、それぞれ、その無音特徴分布を表す正規分布を規定する平均値と分散である。

【0115】

次に、新鮮度関数 $F(x)$ としては、例えば、線形の、離散値をとる関数を用いることができ、この場合、新鮮度関数 $F(x)$ は、例えば、式(48)で表される。

【0116】

【数36】

$$F(x) = \alpha \cdot x \quad x=1, 2, \dots, x_M \quad \dots (48)$$

【0117】

式(48)における α は、所定の定数であり、この定数 α は、式(43)の新鮮度関数の定義から、 $2/(x_M(x_M+1))$ となる。従って、式(48)の新鮮度関数 $F(x)$ は、式(49)で表されることになる。

【0 1 1 8】

【数 3 7】

$$F(x) = \frac{2 \cdot x}{x_M(x_M + 1)} \quad \dots (49)$$

【0 1 1 9】

ここで、式 (49) で表される新鮮度関数 $F(x)$ を、図 1 2 に示す。

【0 1 2 0】

この場合、適応後の無音音響モデル G_s は、式 (50) にしたがって求められる。

【0 1 2 1】

【数 3 8】

$$G_s = \sum_{x=1}^{x_M} \frac{2 \cdot x}{x_M(x_M + 1)} \cdot G_x \quad \dots (50)$$

なお、 G_x は、サンプル点 (サンプル時刻) x における無音特徴分布を表す。

【0 1 2 2】

次に、新鮮度関数 $F(x)$ としては、例えば、指数関数や、高次 2 項関数、対数関数等の非線形の関数を用いることができる。新鮮度関数 $F(x)$ として、例えば、連続値をとる、高次関数としての 2 次関数を用いる場合には、新鮮度関数 $F(x)$ は、例えば、式 (51) で表される。

【0 1 2 3】

【数 3 9】

$$F(x) = \alpha \cdot x^2 \quad \dots (51)$$

【0 1 2 4】

式 (51) における α は、所定の定数であり、この定数 α は、式 (43) の新鮮度関数の定義から、 $3 / x_M^3$ となる。従って、式 (51) の新鮮度関数 $F(x)$ は、式 (52) で表されることになる。

【0 1 2 5】

【数 4 0】

$$F(x) = \frac{3}{x_M^3} \cdot x^2 \quad \dots (52)$$

【0 1 2 6】

ここで、式（52）で表される新鮮度関数 $F(x)$ を、図 1 3 に示す。

【0 1 2 7】

この場合、適応後の無音音響モデル G_s は、式（53）にしたがって求められる。

【0 1 2 8】

【数 4 1】

$$G_s = \frac{3}{x_M^3} \cdot \int_0^{x_M} x^2 \cdot G_x(\mu_x, \sigma_x^2) dx \quad \dots (53)$$

【0 1 2 9】

次に、新鮮度関数 $F(x)$ としては、例えば、離散値をとる、高次関数としての 2 次関数を用いることができ、この場合、新鮮度関数 $F(x)$ は、例えば、式（54）で表される。

【0 1 3 0】

【数 4 2】

$$F(x) = \alpha \cdot x^2 \quad x=1, 2, \dots, x_M \quad \dots (54)$$

【0 1 3 1】

式（54）における α は、所定の定数であり、この定数 α は、式（43）の新鮮度関数の定義から、 $6 / (x_M (x_M + 1) (2x_M + 1))$ となる。従って、式（54）の新鮮度関数 $F(x)$ は、式（55）で表されることになる。

【0 1 3 2】

【数 4 3】

$$F(x) = \frac{6 \cdot x^2}{x_M (x_M + 1) (2x_M + 1)} \quad \dots (55)$$

【0 1 3 3】

ここで、式 (55) で表される新鮮度関数 $F(x)$ を、図 14 に示す。

【0 1 3 4】

この場合、適応後の無音音響モデル G_s は、式 (56) にしたがって求められる。

【0 1 3 5】

【数 4 4】

$$G_s = \sum_{i=1}^{x_M} \frac{6 \cdot x^2}{x_M(x_M+1)(2x_M+1)} \cdot G_x \quad \dots (56)$$

【0 1 3 6】

次に、新鮮度関数 $F(x)$ として、例えば、連続値をとる対数関数を用いる場合には、新鮮度関数 $F(x)$ は、例えば、式 (57) で表される。

【0 1 3 7】

【数 4 5】

$$F(x) = \alpha \cdot \log(x+1) \quad \dots (57)$$

【0 1 3 8】

式 (57) における α は、所定の定数であり、この定数 α は、式 (43) の新鮮度関数の定義から、 $1 / ((x_M+1) \log(x_M+1) - x_M)$ となる。従って、式 (57) の新鮮度関数 $F(x)$ は、式 (58) で表されることになる。

【0 1 3 9】

【数 4 6】

$$F(x) = \frac{1}{(x_M+1) \log(x_M+1) - x_M} \cdot \log(x+1) \quad \dots (58)$$

【0 1 4 0】

ここで、式 (58) で表される新鮮度関数 $F(x)$ を、図 15 に示す。

【0 1 4 1】

この場合、適応後の無音音響モデル G_s は、式 (59) にしたがって求められる。

【0 1 4 2】

【数 4 7】

$$G_s = \frac{1}{(x_M+1) \log(x_M+1) - x_M} \cdot \int_0^{x_M} \log(x+1) \cdot G_x(\mu_x, \sigma_x^2) dx \quad \dots (59)$$

【0 1 4 3】

次に、新鮮度関数 $F(x)$ としては、例えば、離散値をとる対数関数を用いることができ、この場合、新鮮度関数 $F(x)$ は、例えば、式 (60) で表される。

【0 1 4 4】

【数 4 8】

$$F(x) = \alpha \cdot \log(x+1) \quad x=1, 2, \dots, x_M \quad \dots (60)$$

【0 1 4 5】

式 (60) における α は、所定の定数であり、この定数 α は、式 (43) の新鮮度関数の定義から求められ、従って、式 (60) の新鮮度関数 $F(x)$ は、式 (61) で表されることになる。

【0 1 4 6】

【数 4 9】

$$F(x) = \frac{1}{\log \prod_{y=1}^{x_M} (y+1)} \cdot \log(x+1) \quad \dots (61)$$

【0 1 4 7】

ここで、式 (61) で表される新鮮度関数 $F(x)$ を、図 16 に示す。

【0 1 4 8】

この場合、適応後の無音音響モデル G_s は、式 (62) にしたがって求められる。

【0 1 4 9】

【数 5 0】

$$G_s = \frac{1}{\log \prod_{y=1}^{x_M} (y+1)} \cdot \sum_{x=1}^{x_M} \log(x+1) \cdot G_x \quad \dots (62)$$

【0 1 5 0】

次に、新鮮度関数 $F(x)$ として、例えば、連続値をとる、一般的な高次関数を用いる場合には、新鮮度関数 $F(x)$ は、例えば、式 (6 3) で表される。

【0 1 5 1】

【数 5 1】

$$F(x) = \alpha \cdot x^p \quad \dots (63)$$

【0 1 5 2】

式 (6 3) における α は、所定の定数であり、また、 p によって、新鮮度関数 $F(x)$ の次数が決まる。

【0 1 5 3】

定数 α は、式 (4 3) の新鮮度関数の定義から求めることができ、従って、式 (6 3) の新鮮度関数 $F(x)$ は、式 (6 4) で表されることになる。

【0 1 5 4】

【数 5 2】

$$F(x) = \frac{p+1}{x_M^{p+1}} \cdot x^p \quad \dots (64)$$

【0 1 5 5】

この場合、適応後の無音音響モデル G_s は、式 (6 5) にしたがって求められる。

【0 1 5 6】

【数 5 3】

$$G_s = \frac{p+1}{x_M^{p+1}} \cdot \int_0^{x_M} x^p \cdot G_x(\mu_x, \sigma_x^2) dx \quad \dots (65)$$

【0 1 5 7】

なお、式（64）において、例えば、 p が1または2の場合は、新鮮度関数 $F(x)$ は、連続値をとる、線形の関数または2次関数となり、式（46）または（52）に示したように、それぞれ表される。

【0158】

また、式（64）において、例えば、 p が3の場合は、新鮮度関数 $F(x)$ は、連続値をとる3次関数となり、式（66）に示すように表される。

【0159】

【数54】

$$F(x) = \frac{4}{x_M^4} \cdot x^3 \quad \dots (66)$$

【0160】

さらに、式（64）において、例えば、 p が4の場合は、新鮮度関数 $F(x)$ は、連続値をとる4次関数となり、式（67）に示すように表される。

【0161】

【数55】

$$F(x) = \frac{5}{x_M^5} \cdot x^4 \quad \dots (67)$$

【0162】

次に、新鮮度関数 $F(x)$ として、例えば、離散値をとる、一般的な高次関数を用いる場合には、新鮮度関数 $F(x)$ は、例えば、式（68）で表される。

【0163】

【数56】

$$F(x) = \alpha \cdot x^p \quad x=1, 2, \dots, x_M \quad \dots (68)$$

【0164】

式（68）における α は、所定の定数であり、また、 p によって、新鮮度関数 $F(x)$ の次数が決まる。

【0165】

定数 α は、式（43）の新鮮度関数の定義から求めることができ、従って、式

(68) の新鮮度関数 $F(x)$ は、式 (69) で表されることになる。

【0166】

【数57】

$$F(x) = \frac{x^p}{\sum_{y=1}^{x_M} y^p} \quad \dots (69)$$

【0167】

この場合、適応後の無音音響モデル G_s は、式 (70) にしたがって求められる。

【0168】

【数58】

$$G_s = \frac{1}{\sum_{y=1}^{x_M} y^p} \cdot \sum_{x=1}^{x_M} \cdot G_x \quad \dots (70)$$

【0169】

なお、式 (69) において、例えば、 p が 1 または 2 の場合は、新鮮度関数 $F(x)$ は、離散値をとる、線形の関数または 2 次関数となり、式 (49) または (55) に示したように、それぞれ表される。

【0170】

また、式 (69) において、例えば、 p が 3 の場合は、新鮮度関数 $F(x)$ は、離散値値をとる 3 次関数となり、式 (71) に示すように表される。

【0171】

【数59】

$$F(x) = \frac{4x^3}{x_M^2(x_M+1)^2} \quad \dots (71)$$

【0172】

さらに、式 (69) において、例えば、 p が 4 の場合は、新鮮度関数 $F(x)$ は、離散値をとる 4 次関数となり、式 (72) に示すように表される。

【0173】

【数 6 0】

$$F(x) = \frac{4x^3}{x_M(x_M+1)(2x_M+1)(3x_M^2+3x_M-1)} \quad \dots (72)$$

【0 1 7 4】

なお、新鮮度関数 $F(x)$ の概念は、無音音響モデルの適応の他、ノイズ環境下における話者適用や、無音音響モデル以外の音響モデルの適応にも応用することができる。さらに、音声検出や、非定常ノイズ検出にも応用することが可能である。また、音響信号処理や、画像信号処理、通信の分野でも、新鮮度関数 $F(x)$ の概念を用いることで、環境ノイズに対するロバスト性（頑健性）を向上させ、システム性能の改善を図ることが可能となる。

【0 1 7 5】

以上、本発明を適用した音声認識装置について説明したが、このような音声認識装置は、例えば、音声入力可能なカーナビゲーション装置、その他各種の装置に適用可能である。

【0 1 7 6】

なお、本実施の形態では、ノイズの分布特性を考慮した特徴分布パラメータを求めるようにしたが、このノイズには、例えば、発話を行う環境下における外部からのノイズの他、例えば、電話回線その他の通信回線を介して送信されてくる音声の認識を行う場合には、その通信回線の特性なども含まれる。

【0 1 7 7】

また、本発明は、音声認識の他、画像認識その他のパターン認識を行う場合にも適用可能である。

【0 1 7 8】

さらに、本実施の形態では、特徴空間における分布として表される無音特徴分布を用いて、無音音響モデルの適応を行うようにしたが、無音音響モデルの適応は、特徴空間における点として表されるノイズの特徴量を用いて行うことも可能である。

【0 1 7 9】

次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフ

トウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0 1 8 0】

そこで、図 1 7 は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0 1 8 1】

プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク 1 0 5 や ROM 1 0 3 に予め記録しておくことができる。

【0 1 8 2】

あるいはまた、プログラムは、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体 1 1 1 に、一時的あるいは永続的に格納(記録)しておくことができる。このようなリムーバブル記録媒体 1 1 1 は、いわゆるパッケージソフトウェアとして提供することができる。

【0 1 8 3】

なお、プログラムは、上述したようなリムーバブル記録媒体 1 1 1 からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部 1 0 8 で受信し、内蔵するハードディスク 1 0 5 にインストールすることができる。

【0 1 8 4】

コンピュータは、CPU(Central Processing Unit) 1 0 2 を内蔵している。CPU 1 0 2 には、バス 1 0 1 を介して、入出力インタフェース 1 1 0 が接続されており、CPU 1 0 2 は、入出力インタフェース 1 1 0 を介して、ユーザによって、キーボードやマウス等で構成される入力部 1 0 7 が操作されることにより指令が入

力されると、それにしたがって、ROM(Read Only Memory) 103に格納されているプログラムを実行する。あるいは、また、CPU 102は、ハードディスク 105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部 108で受信されてハードディスク 105にインストールされたプログラム、またはドライブ 109に装着されたりムーバブル記録媒体 111から読み出されてハードディスク 105にインストールされたプログラムを、RAM(Random Access Memory) 104にロードして実行する。これにより、CPU 102は、上述したブロック図の構成により行われる処理を行う。そして、CPU 102は、その処理結果を、必要に応じて、例えば、入出力インタフェース 110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部 106から出力、あるいは、通信部 108から送信、さらには、ハードディスク 105に記録等させる。

【0185】

ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0186】

また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0187】

【発明の効果】

本発明のモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置によれば、所定区間における抽出データと、その抽出データの新鮮さを表す新鮮度に基づいて、所定のモデルの適応が行われる。従って、そのモデルを用いてパターン認識を行うことで、認識性能を向上させることが可能となる。

【図面の簡単な説明】

【図1】

本発明を適用した音声認識装置の構成例を示すブロック図である。

【図 2】

図 1 のノイズ観測区間抽出部 3 の動作を説明するための図である。

【図 3】

図 1 の特徴抽出部 5 の詳細な構成例を示すブロック図である。

【図 4】

図 1 の音声認識部 6 の詳細な構成例を示すブロック図である。

【図 5】

HMMを示す図である。

【図 6】

シミュレーション結果を示す図である。

【図 7】

無音音響モデルの正規分布を示す図である。

【図 8】

図 1 の無音音響モデル補正部 7 の構成例を示すブロック図である。

【図 9】

離散値を連続値に変換する様子を示す図である。

【図 1 0】

一般的な新鮮度関数 $F(x)$ を示す図である。

【図 1 1】

新鮮度関数 $F(x)$ の第 1 の例を示す図である。

【図 1 2】

新鮮度関数 $F(x)$ の第 2 の例を示す図である。

【図 1 3】

新鮮度関数 $F(x)$ の第 3 の例を示す図である。

【図 1 4】

新鮮度関数 $F(x)$ の第 4 の例を示す図である。

【図 1 5】

新鮮度関数 $F(x)$ の第 5 の例を示す図である。

【図 1 6】

新鮮度関数 $F(x)$ の第 6 の例を示す図である。

【図 1 7】

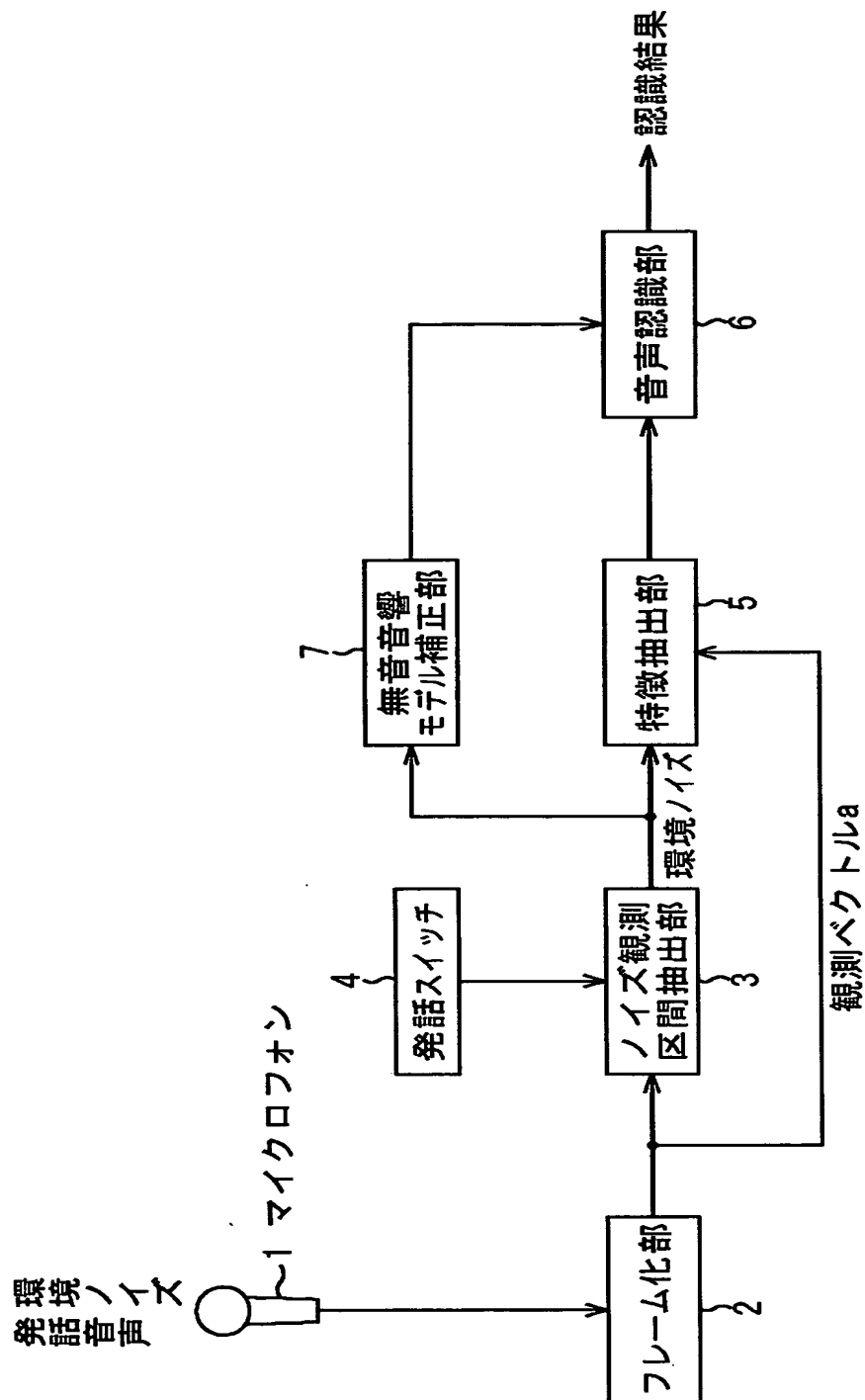
本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

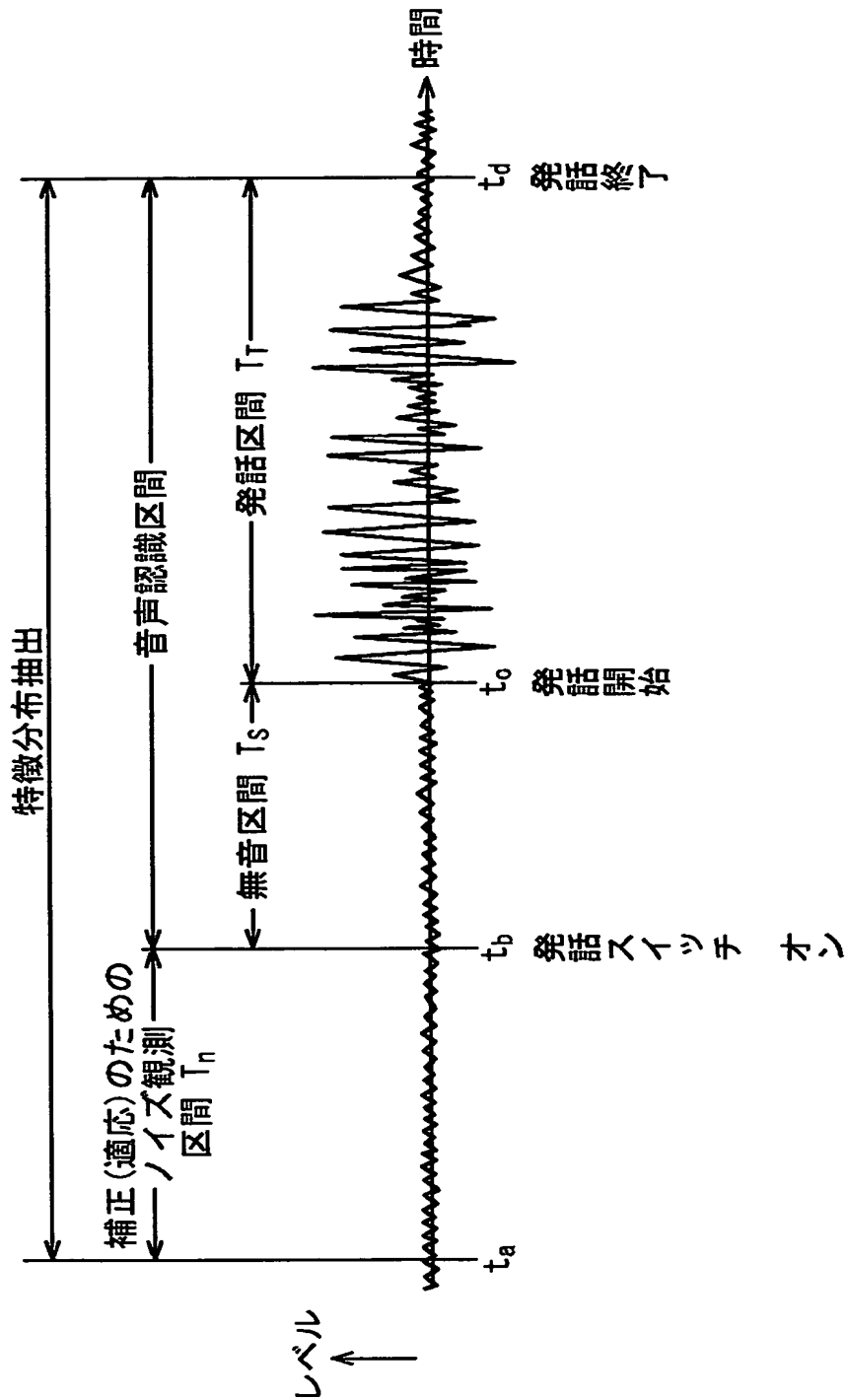
1 マイクロフォン, 2 フレーム化部, 3 ノイズ観測区間抽出部,
4 発話スイッチ, 5 特徴抽出部, 6 音声認識部, 7 無音音響モデル補正部, 31 新鮮度関数記憶部, 32 補正部, 101 バス, 102 CPU, 103 ROM, 104 RAM, 105 ハードディスク, 106 出力部, 107 入力部, 108 通信部, 109 ドライブ, 110 入出力インタフェース, 111 リムーバブル記録媒体

【書類名】 図面

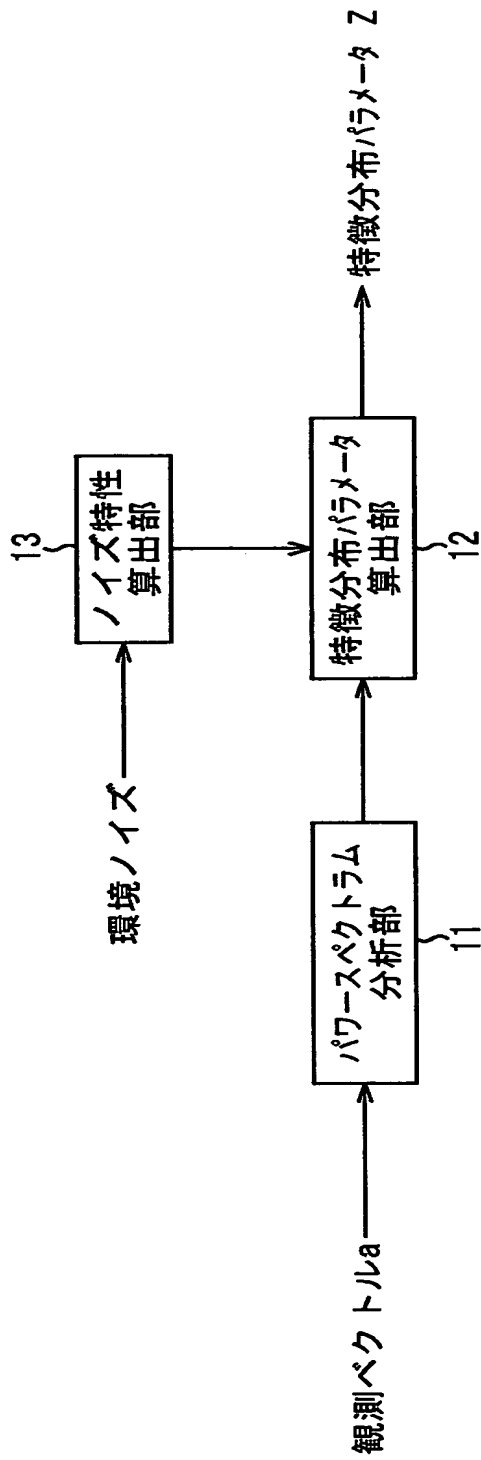
【図 1】



【図 2】

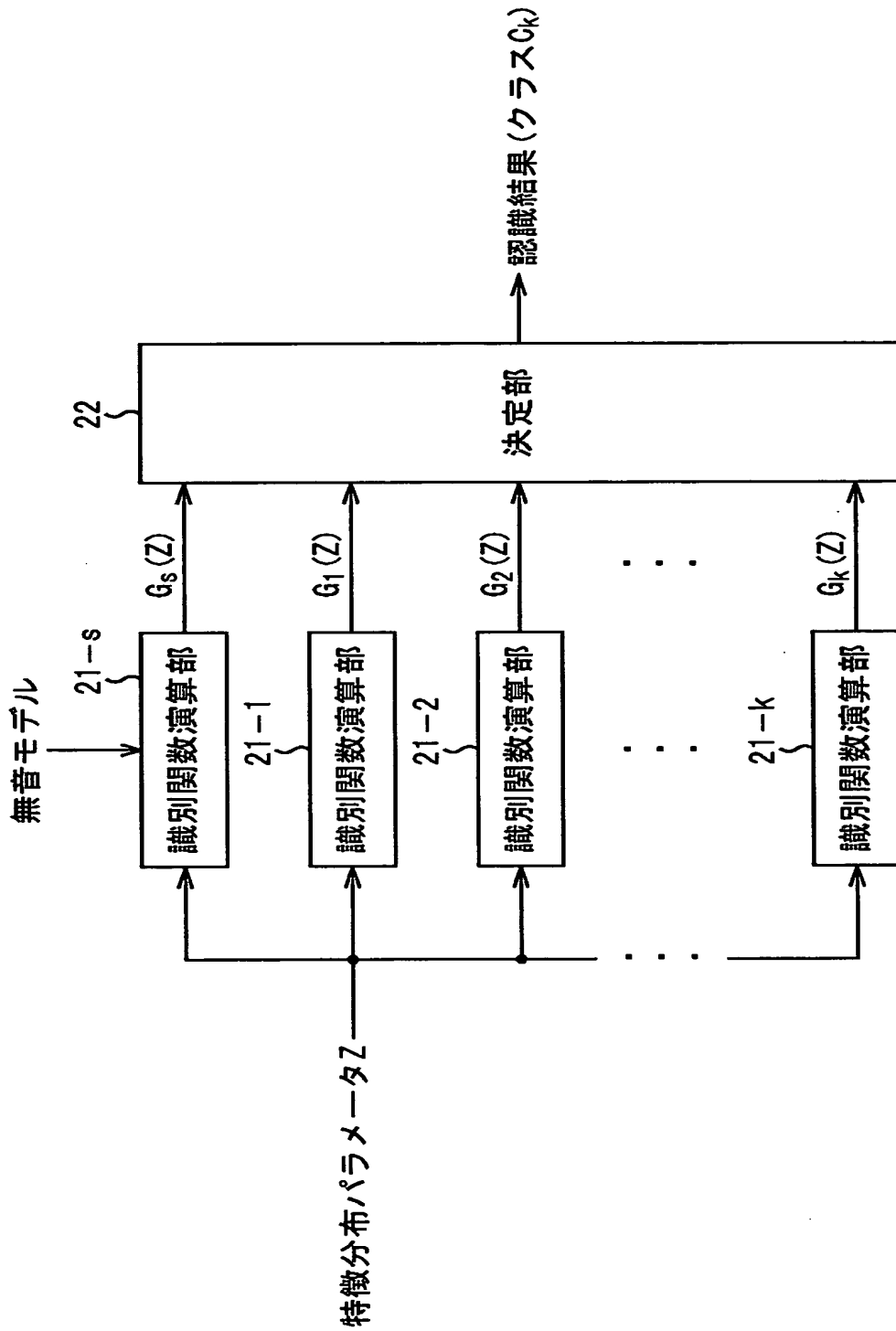


【図 3】



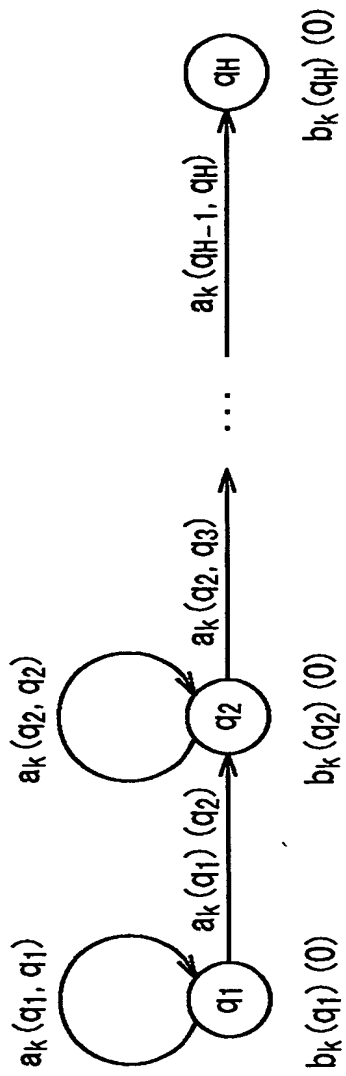
特徴抽出部 5

【図 4】



音声認識部 6

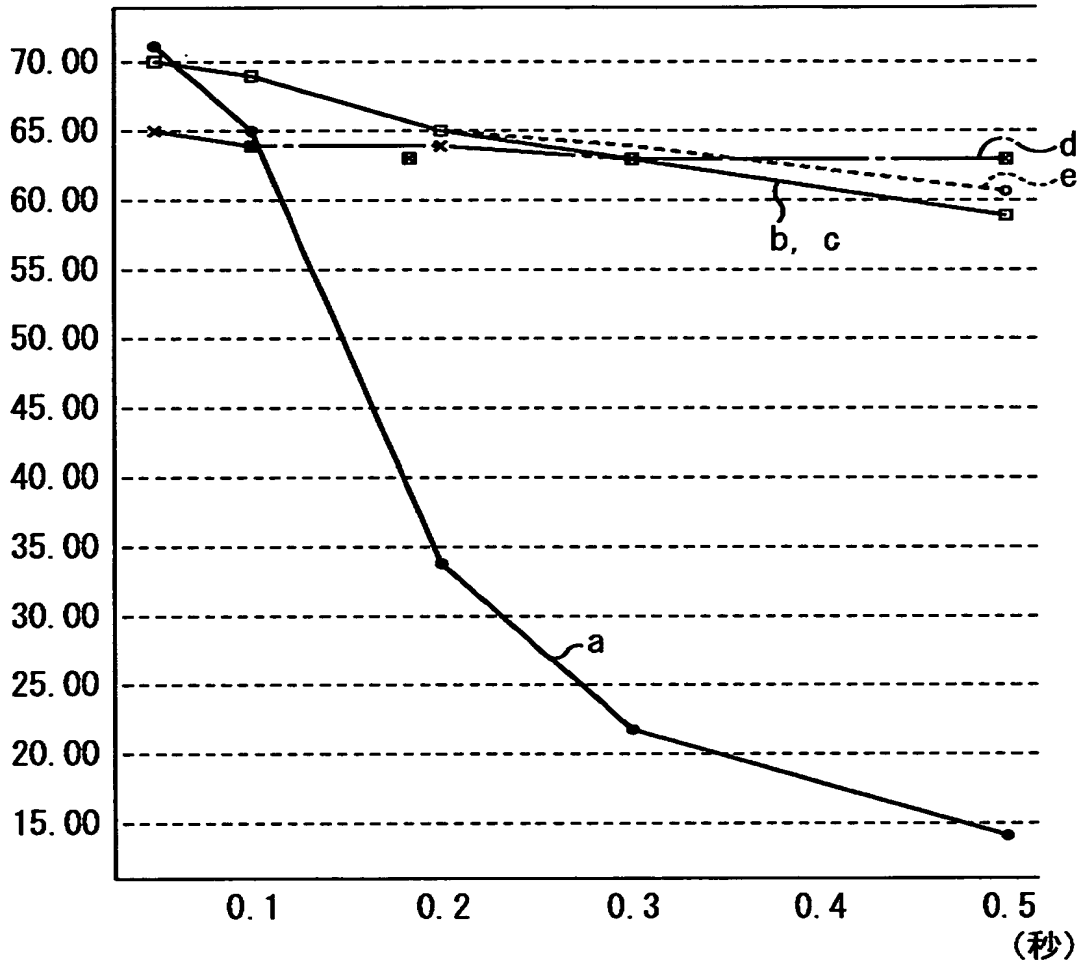
【図 5】



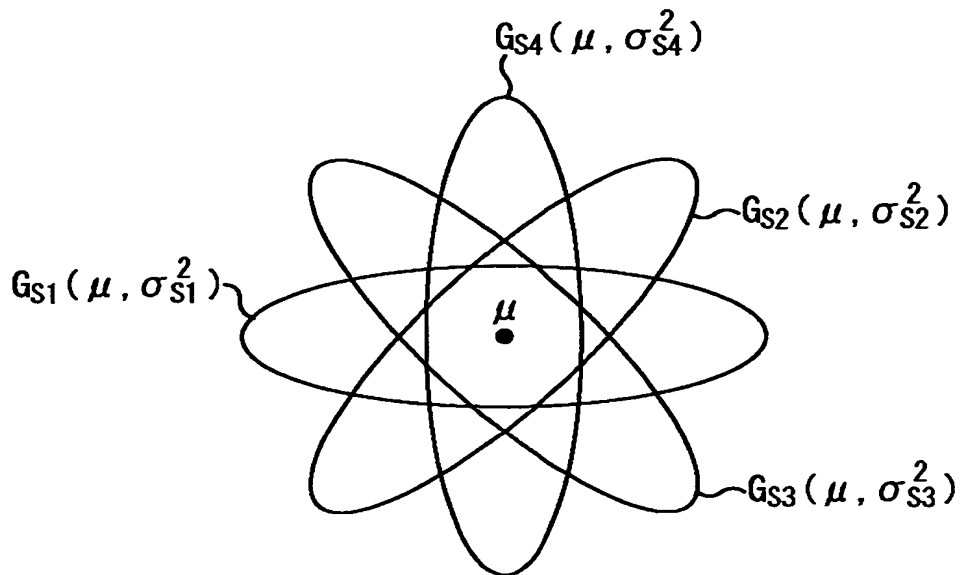
HMM (left-to-right モデル)

【図 6】

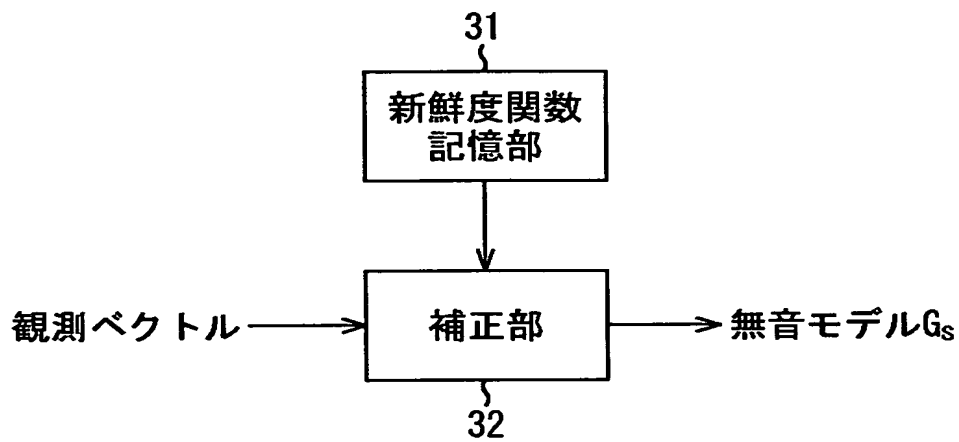
認識率 (%)



【図 7】

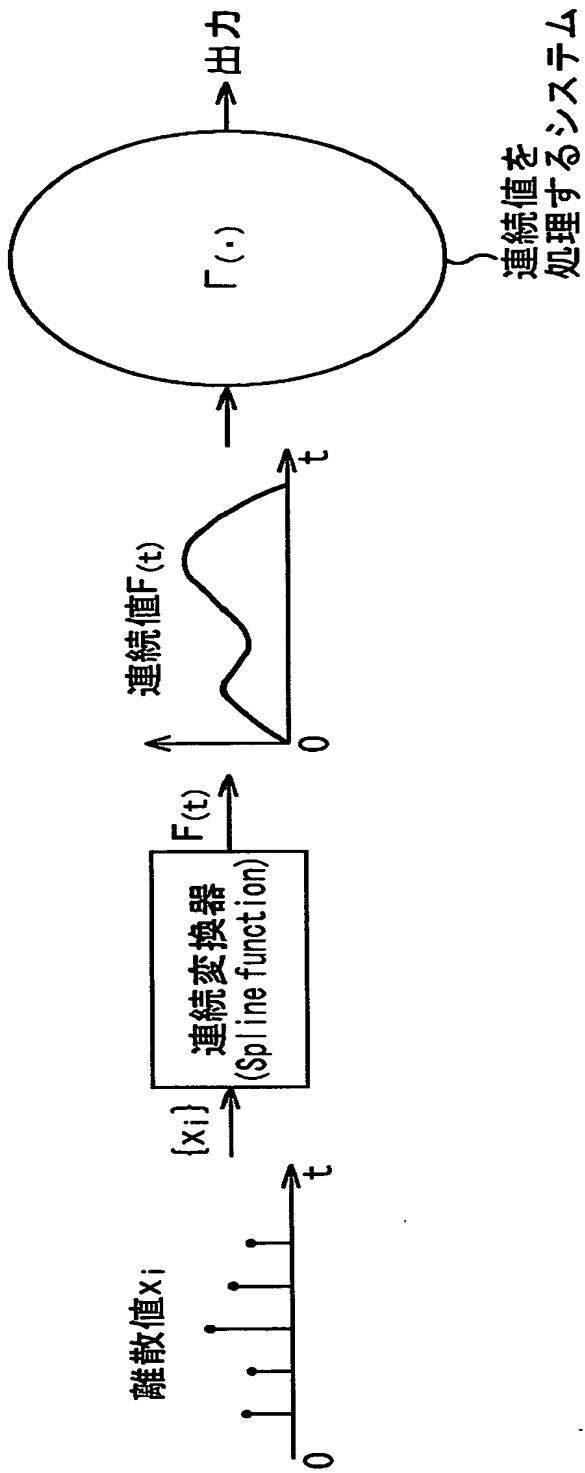


【図 8】

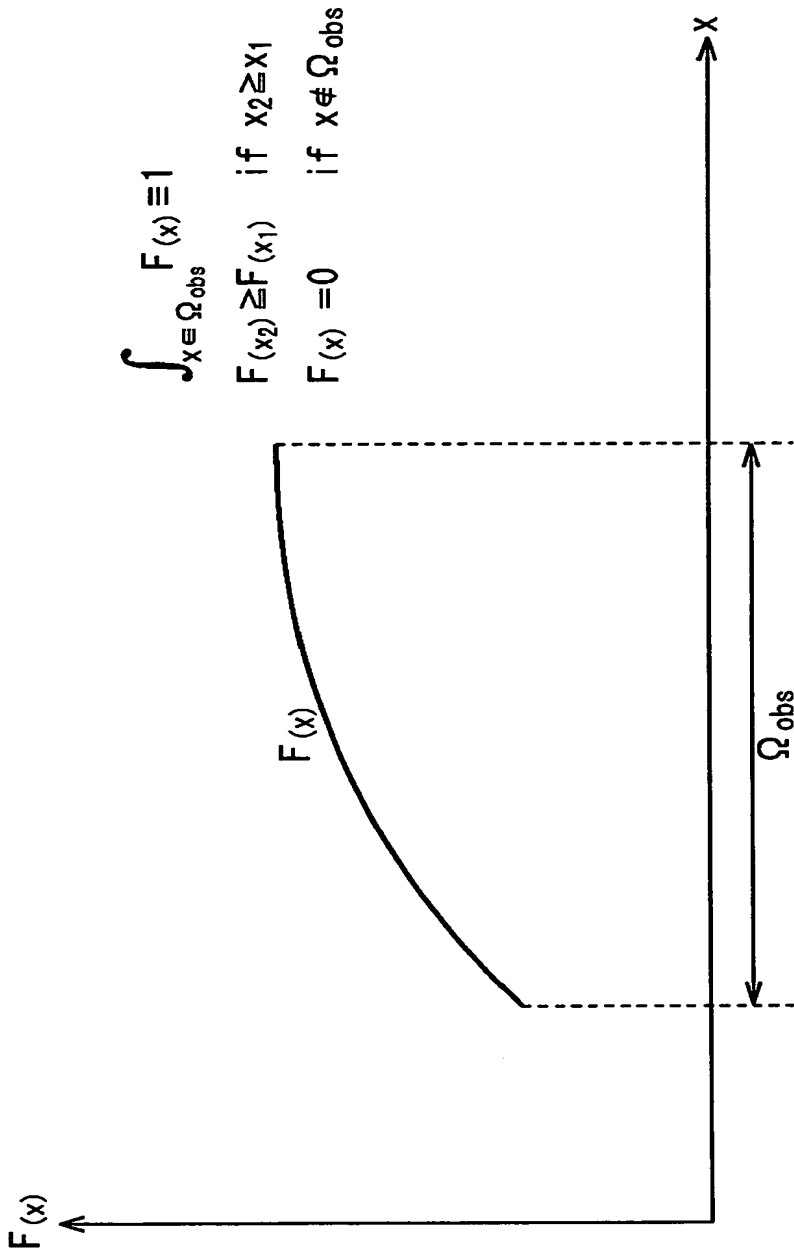


無音音響モデル補正部 7

【図 9】

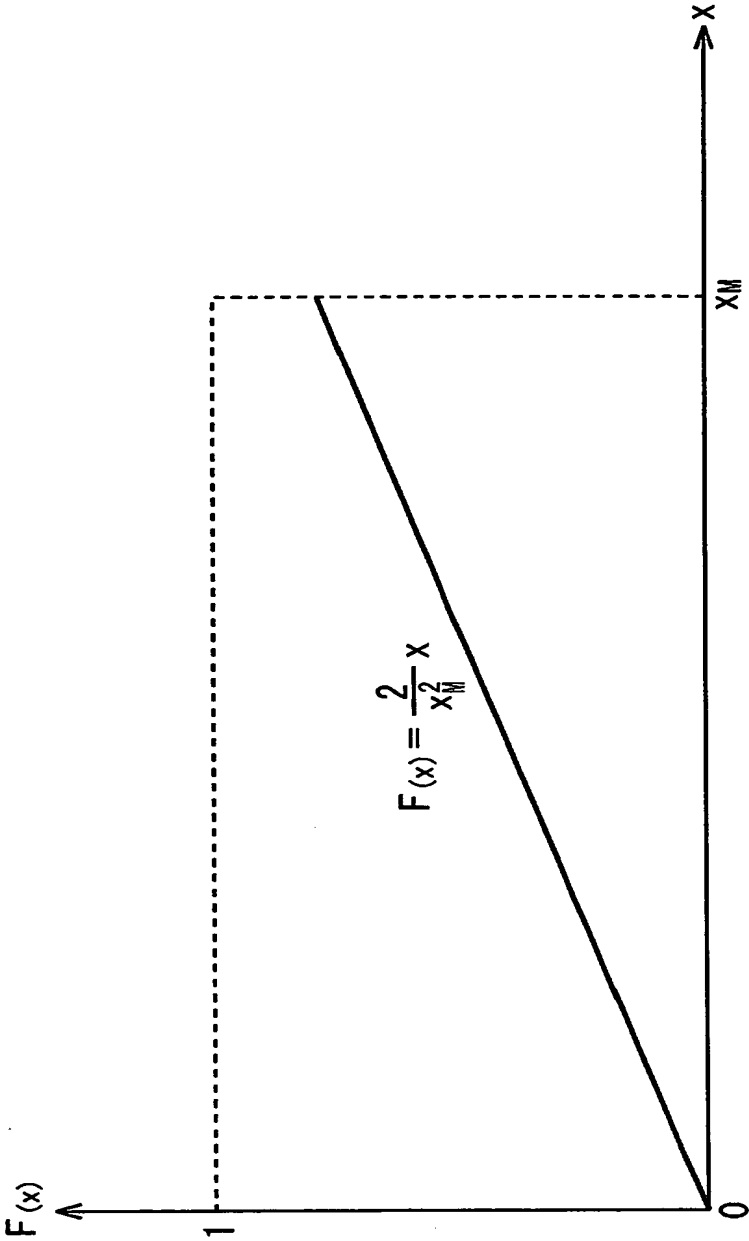


【図 1 0】



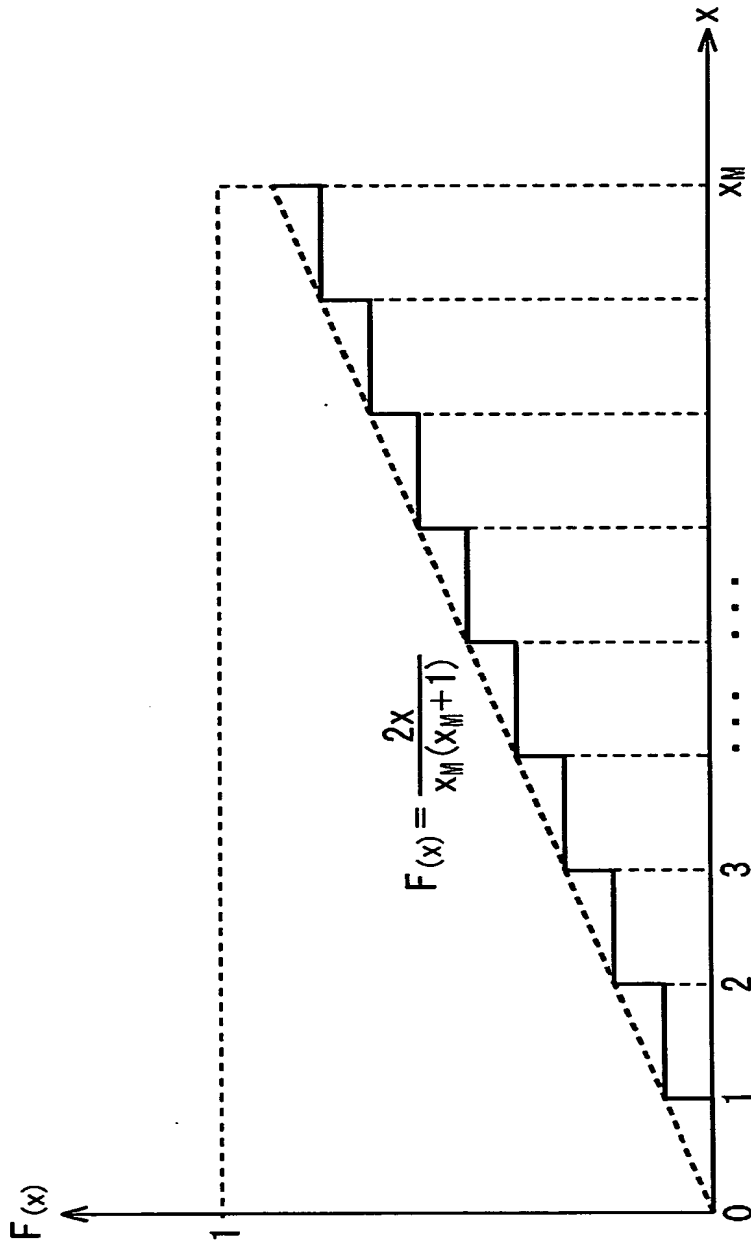
連続線形新鮮度関数 (Continuous Linear Freshness Function)

【図 1 1】



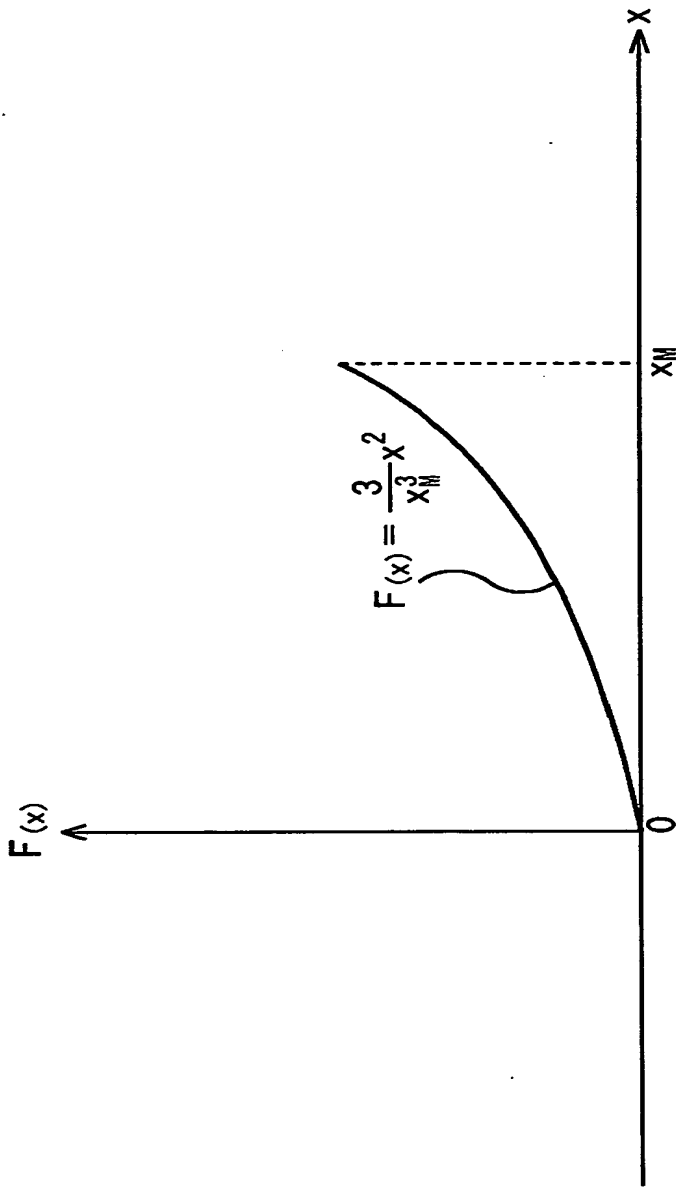
【図 1 2】

離散線形新鮮度関数 (Discrete Linear Freshness Function)



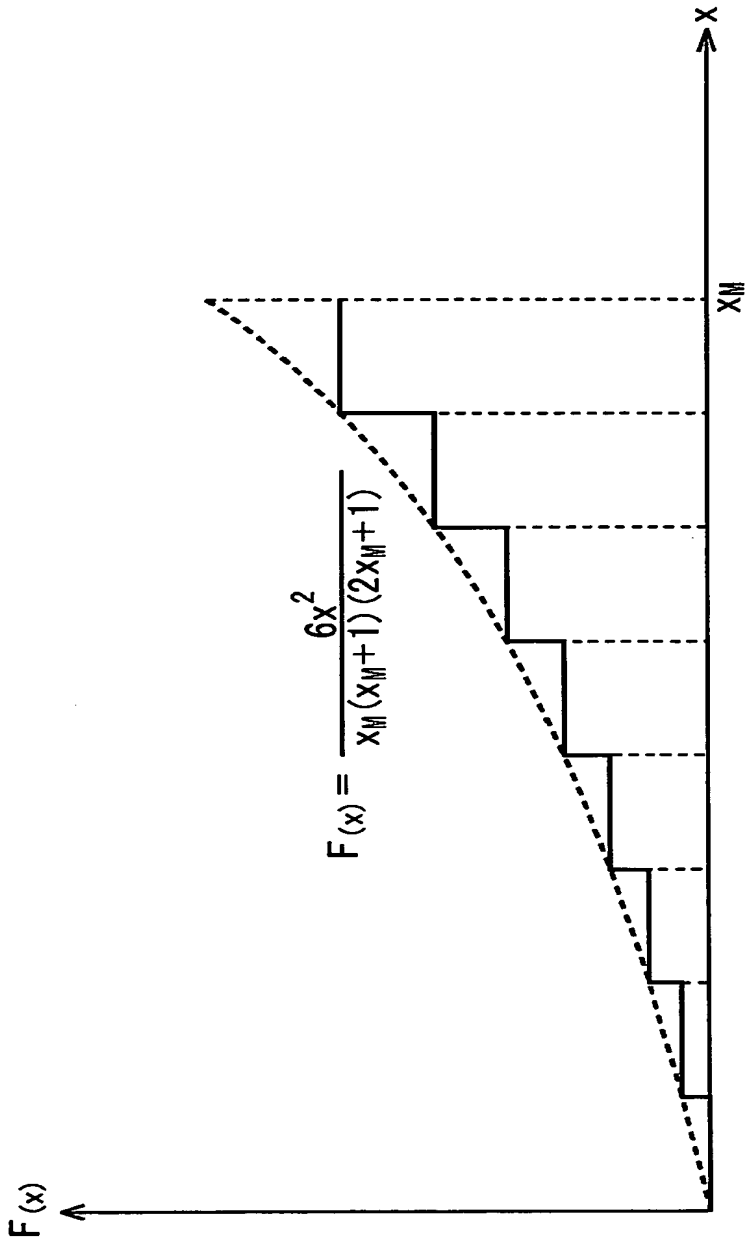
連続非線形2次新鮮度関数 (2-order Continuous Freshness Function)

【図 1 3】



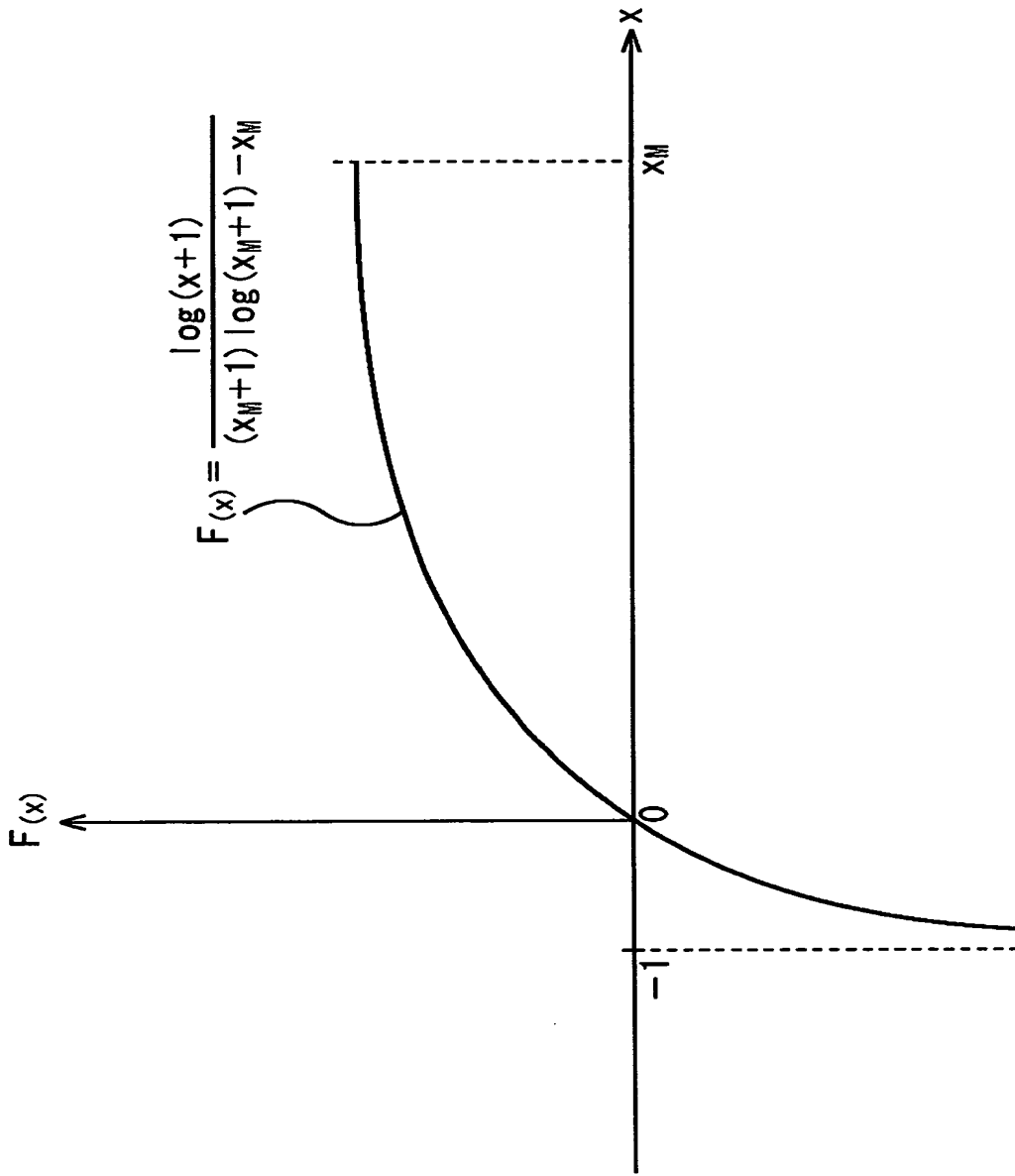
【図 1 4】

離散非線形2次新鮮度関数 (2-order Discrete Freshness Function)



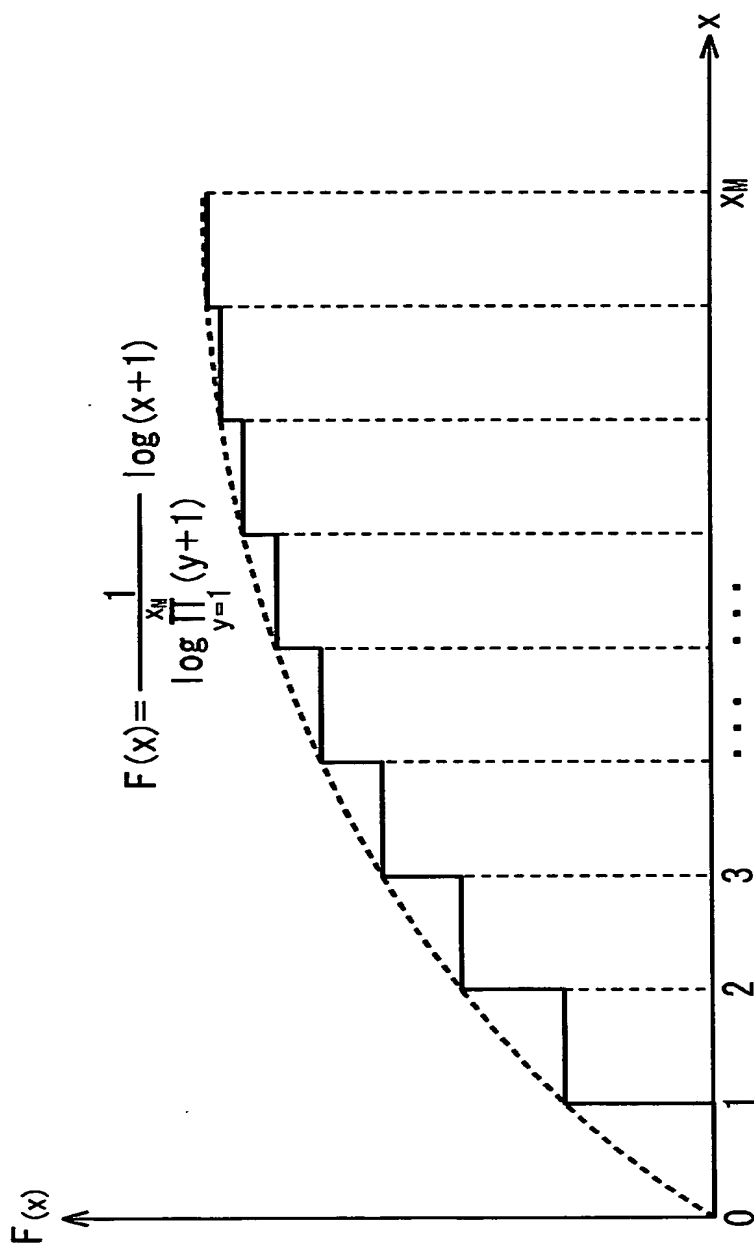
【図 1 5】

連続非線形対数新鮮度関数 (Log-order Continuous Freshness Function)

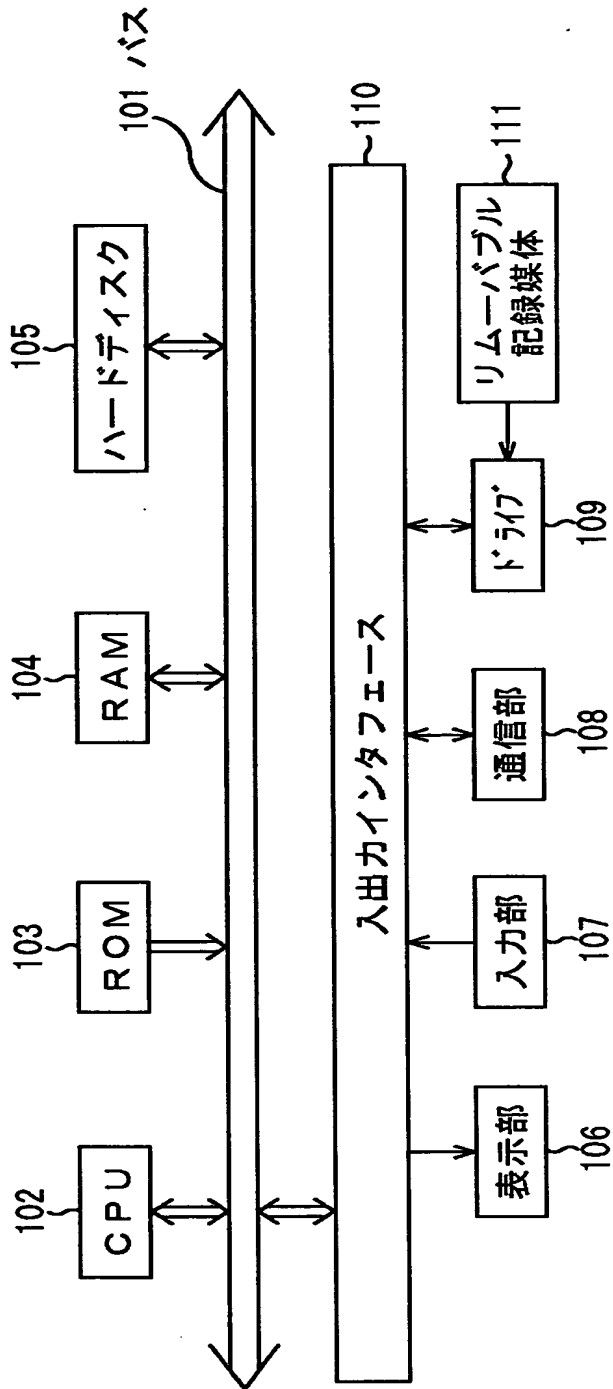


【図 1 6】

離散非線形対数新鮮度関数 (Log-order Discrete Freshness Function)



【図 1 7】



コンピュータ

【書類名】 要約書

【要約】

【課題】 認識性能を向上させる。

【解決手段】 無音音響モデル補正部 7 は、無音の状態を表す音響モデルである無音音響モデルの適応を、音声認識の対象とする音声認識区間の直前の区間で観測される入力データと、その入力データの新鮮さを表す新鮮度に基づいて行う。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 2 1 8 5]

1. 変更年月日 1 9 9 0 年 8 月 3 0 日

[変更理由] 新規登録

住 所 東京都品川区北品川 6 丁目 7 番 3 5 号

氏 名 ソニー株式会社